

## ARTICLE OPEN



# Ubiquitous, B<sub>12</sub>-dependent viroplankton utilizing ribonucleotide-triphosphate reductase demonstrate interseasonal dynamics and associate with a diverse range of bacterial hosts in the pelagic ocean

Ling-Yi Wu<sup>1</sup>, Gonçalo J. Piedade<sup>2,3</sup>, Ryan M. Moore<sup>4</sup>, Amelia O. Harrison<sup>4</sup>, Ana M. Martins<sup>3</sup>, Kay D. Bidle<sup>5</sup>, Shawn W. Polson<sup>4</sup>, Eric G. Sakowski<sup>6</sup>, Jozef I. Nissimov<sup>7</sup>, Jacob T. Dums<sup>4,8</sup>, Barbra D. Ferrell<sup>4</sup> and K. Eric Wommack<sup>4</sup>✉

© The Author(s) 2023

Through infection and lysis of their coexisting bacterial hosts, viruses impact the biogeochemical cycles sustaining globally significant pelagic oceanic ecosystems. Currently, little is known of the ecological interactions between lytic viruses and their bacterial hosts underlying these biogeochemical impacts at ecosystem scales. This study focused on populations of lytic viruses carrying the B<sub>12</sub>-dependent Class II monomeric ribonucleotide reductase (RNR) gene, ribonucleotide-triphosphate reductase (Class II RTPR), documenting seasonal changes in pelagic viroplankton and bacterioplankton using amplicon sequences of Class II RTPR and the 16S rRNA gene, respectively. Amplicon sequence libraries were analyzed using compositional data analysis tools that account for the compositional nature of these data. Both virio- and bacterioplankton communities responded to environmental changes typically seen across seasonal cycles as well as shorter term upwelling–downwelling events. Defining Class II RTPR-carrying viral populations according to major phylogenetic clades proved a more robust means of exploring viroplankton ecology than operational taxonomic units defined by percent sequence homology. Viroplankton Class II RTPR populations showed positive associations with a broad phylogenetic diversity of bacterioplankton including dominant taxa within pelagic oceanic ecosystems such as *Prochlorococcus* and SAR11. Temporal changes in Class II RTPR viroplankton, occurring as both free viruses and within infected cells, indicated possible viral–host pairs undergoing sustained infection and lysis cycles throughout the seasonal study. Phylogenetic relationships inferred from Class II RTPR sequences mirrored ecological patterns in virio- and bacterioplankton populations demonstrating possible genome to phenome associations for an essential viral replication gene.

ISME Communications; <https://doi.org/10.1038/s43705-023-00306-9>

## INTRODUCTION

The subtropical pelagic ocean gyres are extensive, coherent regions that collectively represent the largest ecosystems on Earth, accounting for 40% of the earth's surface [1]. Circulation within the upper kilometer of subtropical gyres is primarily wind driven [2]. Due to their size, the horizontal and vertical motion of water within this upper layer of the gyres plays an outsized role in regulating global nutrient cycles and the size of the atmospheric carbon pool. Both issues are critical concerns for understanding anthropogenic impacts on fisheries, oceanic general circulation, and global climate. Ultimately, the activities of and the interactions between oceanic microbial communities—composed of protists, phytoplankton, bacterioplankton, and viroplankton—regulate the flow of carbon and nutrient elements within the

pelagic ocean gyres [3, 4]. Truly appreciating how oceans will respond to environmental changes brought on by anthropogenic activities requires detailed, mechanistic understanding of oceanic microbial communities [5–8]. This study focused on specific interactions between important and ubiquitous populations of lytic viruses [9, 10] and bacterioplankton, the most abundant microbial host group within the low nutrient ecosystems that dominate much of the pelagic ocean.

Viral cell lysis is among the microbial community interactions most directly responsible for the conversion of cellular biomass into dissolved and particulate organic carbon [11]. This process is one of the major top-down controls on bacterioplankton productivity, responsible for between 10 and 40% of daily bacterioplankton mortality in the ocean [12]. Upon lysis, cell

<sup>1</sup>Theoretical Biology and Bioinformatics, Science4Life, Utrecht University, Padualaan 8, Utrecht 3584 CH, the Netherlands. <sup>2</sup>Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, 1797 SZ t'Horntje, The Netherlands. <sup>3</sup>Department of Oceanography and Fisheries and Ocean Sciences Institute-OKEANOS, University of the Azores, 9901-862 Horta, Faial, Azores, Portugal. <sup>4</sup>Delaware Biotechnology Institute, University of Delaware, 590 Avenue 1743, Newark, DE 19713, USA. <sup>5</sup>Department of Marine and Coastal Sciences, Rutgers University, 71 Dudley Rd., New Brunswick, NJ 08901, USA. <sup>6</sup>Department of Earth Sciences, Johns Hopkins University, Baltimore, MD, USA. <sup>7</sup>Department of Biology, University of Waterloo, 200 University Ave. West, Waterloo, ON N2L 3G1, Canada. <sup>8</sup>Biotechnology Program, North Carolina State University, 2800 Faucette Dr, Raleigh, NC 27695, USA. ✉email: [wommack@udel.edu](mailto:wommack@udel.edu)

Received: 30 January 2023 Revised: 31 August 2023 Accepted: 6 September 2023

Published online: 03 October 2023

contents not consumed by viral replication are released, contributing to dissolved nutrient pools within marine ecosystems. Bacterioplankton cell lysis comprises the largest known carbon flux process between biomass and the dissolved organic matter (DOM) pool amounting to ~150 gigatons C/yr [6, 13]. The conversion of biomass to DOM from viral lysis is five times greater than that performed by other biological mechanisms such as grazing and programmed cell death [6, 13]. This flux from biomass to DOM sustains nutrient cycles, providing essential growth substrates for bacterioplankton and phytoplankton productivity. The availability and concentrations of different nutrients within the DOM pool comprise the bottom-up controls on bacterioplankton productivity [6, 13, 14]. Besides influencing the composition, diversity, and productivity of marine bacterial communities through top-down and bottom-up controls, viruses can influence bacterial physiology, improving fitness in the face of environmental changes [15, 16] by altering metabolic pathways [17, 18] through horizontal gene transfer; and through the expression of metabolic genes carried within viral genomes during viral infection [19, 20]. However, the behavior of individual viroplankton populations within ecosystems and their detailed interactions with bacterioplankton host populations remain largely unknown. This study examined both viroplankton and bacterioplankton dynamics within the North Atlantic Subtropical Gyre across time with the objective of observing interactions at the community and population levels.

For some time, microbial ecologists have investigated bacterial populations and communities utilizing 16S rRNA gene amplicon sequences, a universal gene marker for bacterial diversity [21]. Due to their polyphyletic origins, viruses do not have a universally conserved marker gene that could be similarly used for tracing viral evolutionary history and community ecology investigations. Nevertheless, there are a few widespread genes exhibiting robust evolutionary histories that can be used as markers for viral community ecology studies within natural environments. Genes such as the DNA polymerase A gene (encoding DNA polymerase I) [22, 23], terminase gp20 gene [24–27], major capsid protein gp23 gene of T4-like phages [28–30], and ribonucleotide reductase (RNR) genes [9] have all informed studies assessing the diversity of abundant viral populations. This study used amplicons of the ribonucleotide-triphosphate reductase (Class II RTPR) gene (*nrdJm*) for following the dynamics of viroplankton populations.

RNRs are the only known enzymes capable of reducing ribonucleotides to deoxyribonucleotides, essential substrates for DNA synthesis [31, 32]. RNRs are ancient enzymes that were essential to the emergence of DNA [33] and all cellular life and all DNA viruses require this enzyme for genome replication. As a consequence, RNR genes are under rigorous evolutionary selection pressure and among the most abundantly identified genes in marine virome libraries [34]. Importantly, RNR genes are present in tailed phages within the class Caudoviricetes and have been identified within the genomes of viruses infecting hosts within all three domains of life. Generally, lytic marine viruses carry RNR genes [35], which ensure a steady supply of deoxyribonucleotides for DNA synthesis [36–38]. RNRs are biologically informative as this group of enzymes consists of three physiological classes according to reactivity with oxygen. Class I RNRs are oxygen-dependent. Class II RNRs are oxygen-independent and rely upon an adenosylcobalamin (vitamin B<sub>12</sub>) co-factor for ribonucleotide reduction. Class II RNRs are further divided according to structure and substrate utilization. Monomeric Class II RTPR reduces ribonucleotide triphosphate to deoxyribonucleotide triphosphate (dNTPs), the substrate for DNA synthesis. Dimeric Class II reduces ribonucleotide diphosphate to deoxyribonucleotide diphosphate thus requiring a dinucleotide kinase for the final step in forming dNTPs. Class III RNRs are sensitive to oxygen. As such, Class I and Class II RNRs, not Class III RNRs, are commonly present within viroplankton communities found in oxygenated ocean water [10, 32].

The frequency of RNR types varies substantially among known cellular organisms and viruses (Supplementary Fig. S1). Meta-analysis of RNR protein sequences within the RNRdb [39] along with sequences in the UniProt knowledgebase [40] showed the high frequency of Class II RTPR (i.e., Class II monomeric RNR) within known viruses (~22%) as compared to cellular kingdoms (2–3%). Within the Caudovirales most Class II RTPR genes occurred within siphophages (Supplementary Fig. S1). In contrast, Archaea and Bacteria favor Class II dimeric or Class III RNRs for anaerobic ribonucleotide reduction, a biochemical strategy strikingly different from viruses. Given the high frequency and unique biochemical features of Class II RTPRs, we hypothesized that viral populations carrying this gene would interact with a cross-section of bacterioplankton taxa and demonstrate dynamic behavior befitting lytic viruses reliant on actively growing bacterioplankton host populations.

This study leveraged the predominance of Class II RTPR genes within viruses [9] for examining the ecological dynamics of viroplankton populations. Class II RTPR and 16S rRNA amplicon sequence data were analyzed against a backdrop of microbial and oceanographic observations using the latest compositional data analysis (CoDA) tools [41–46] for assessing how lytic, B<sub>12</sub>-dependent viral populations responded to seasonal dynamic changes within the euphotic zone of a subtropical oceanic gyre, a region critically important within the global carbon cycle.

## MATERIALS AND METHODS

### Study site and sample collection

Thirty-five seawater samples were collected across seven sampling dates (May–September, separated by three to four weeks) and five depths from a fixed station at South of Channel Faial-Pico islands (SCFP; 38° 23′ 11.67″ N, 28° 35′ 6.14″ W), located in the northeastern Atlantic, 6.4 nautical miles off São Mateus (Pico Island) in the Azores archipelago (ICES subdivision Xa2). The station is representative of a larger open ocean ecosystem, the Mid-Atlantic Ridge ecoregion, corresponding to the International Council for the Exploration of the Seas (ICES) areas X and XII, which includes a high proportion of high seas [47]. The Azores are located along the northern boundary of the North Atlantic Subtropical Gyre, which constitutes a barrier between superficial cold waters of Nordic provenance and warmer waters of the gyre. Due to the influence of large-scale currents such as the Gulf Stream, the North Atlantic Current, and the Azores Current, the islands are mostly exposed to the transport of warm waters of equatorial and tropical origin. These currents result in high salinity, sharp horizontal temperature gradients, and high sea surface temperatures (on average 15–16°C in winter and 22–24°C in summer). These physical conditions result in low nutrient regimes and correspond with low ecosystem productivity [48]. Nevertheless, frontal regions and underwater features (seamounts, eddies, upwelling areas) can enhance local productivity [49, 50].

Water samples were collected in acid-rinsed (0.1 M HCl) and seawater-flushed carboys between 9:30 and 11:00 local time from 0, 5, 25, 50, and 75 m depths during seven sampling events on May 12th, June 8th, June 22nd, July 11th, July 27th, August 5th, and September 8th of 2016. Surface seawater (0 m depth) was collected by directly submerging a carboy. Seawater at other depths was collected using a 2.5 L Niskin bottle and subsequently transferred to a 5 L carboy. Seawater samples were transported inside a light-tight cooler and laboratory-processed within 2 h of collection. Subsamples for flow cytometric enumeration of phytoplankton were fixed immediately using glutaraldehyde (0.5% final concentration). Flow cytometry samples were transported on ice (1–2 h) and stored at –80 °C until further processing.

Water column conductivity, temperature, and depth (CTD) data was collected down to 150 m using a CTD instrument (MIDAS Valeport, Totnes, United Kingdom) (Supplementary Table S1). Salinity (unitless, on the Practical Salinity Scale PSS-78) was expressed as a function of conductivity. Density (kg/m<sup>3</sup>) was expressed as a function of pressure. CTD profiles were visualized using the cmocean color palettes [51] on Ocean Data View 5.3.0 [52] (<https://odv.awi.de>). Temperature and salinity values between sampling days were interpolated using the weighted average grid method [52]. Temperature, salinity, and density were determined by averaging all values 0.5 m above and below the targeted sampling depths. Equipment

malfunction prevented CTD data collection on September 8th, thus, temperature, salinity, and density metadata were estimated from the nearest time, pixel, and depth from the Marine Copernicus Global Ocean Physics model [53] (GLOBAL\_ANALYSISFORECAST\_PHY\_CPL\_001\_015, <http://marine.copernicus.eu/>). Average values for chlorophyll *a* ( $\text{mg}/\text{m}^3$ ), sea surface temperature (SST,  $^{\circ}\text{C}$ ), particulate organic carbon (POC,  $\text{mg}/\text{m}^3$ ), and photosynthetically active radiation (PAR,  $\text{Einstein}/\text{m}^2\text{d}$ ) data in 2016 were obtained from the NASA Ocean Color website (<http://oceancolor.gsfc.nasa.gov/>) at 1 km resolution. Satellite data was mapped using SeaDAS 7 [54] for the  $9\text{ km}^2$  area surrounding sampling station SCFP.

### Phytoplankton Enumeration using Flow Cytometry

Phytoplankton cells were enumerated using a BD Influx Mariner 209 s Flow Cytometer (BD Biosciences, San Jose, CA) equipped with a 488 nm argon laser and  $525 \pm 15$ ,  $542 \pm 27$ ,  $585 \pm 40$ , and  $692 \pm 40$  nm emission filters with a nozzle size of 70  $\mu\text{m}$ . After thawing and gentle vortexing, samples were run for 2 min [55]. A blank consisting of 0.22  $\mu\text{m}$  filtered Milli-Q water was run before and after counting. The flow rate was determined by running Milli-Q water for 2 min and dividing the mass difference (determined with a scale) by time; mass was converted to volume using the density of 1  $\text{kg}/\text{L}$ . Phytoplankton cells were identified based on chlorophyll fluorescence ( $692 \pm 40$  nm) from 488 nm excitation (Supplementary Table S1).

### Virio- and bacterioplankton biomass collection

The cellular ( $>0.22\ \mu\text{m}$ ) and subcellular (0.02–0.22  $\mu\text{m}$ ) fractions, biologically defined in this study as bacterioplankton and viroplankton, respectively, were separated through two consecutive syringe filtrations using 0.22 and 0.02  $\mu\text{m}$  filters. All tubing and containers were acid-washed and deionized water-rinsed prior to sample filtration. The bacterioplankton fraction was collected by filtering 5 L of seawater through a 0.22  $\mu\text{m}$  pore-size Sterivex filter (Merck Millipore, Darmstadt, Germany) under a vacuum pressure system linked to a filtration ramp (Merck Millipore, Burlington, MA). The sample collection flask was rinsed twice using the first 200 mL of 0.22  $\mu\text{m}$  filtrate. Subsequently, the viroplankton fraction was collected by filtering 1 L of 0.22  $\mu\text{m}$  filtrate through a 0.02  $\mu\text{m}$  pore-size Anotop filter (Whatman, Maidstone, UK) [56, 57] connected to a 60 mL syringe, using a caulk gun which generated a steady pressure on the syringe. Both 0.22  $\mu\text{m}$  and 0.02  $\mu\text{m}$  filters with bacterio- or viroplankton biomass were parafilm sealed and stored at  $-80\ ^{\circ}\text{C}$  until DNA extraction.

### DNA Extraction from the Viroplankton Fraction and Tag-Encoded Class II RTPR Amplicon Sequencing

In brief, DNA was extracted from the 0.02  $\mu\text{m}$  Anotop filters (viroplankton fraction) and quantified using previously reported protocols [57]. Class II RTPR amplicons of  $\sim 750$  bp were obtained using viroplankton template DNA and degenerate primers designed using sequence alignments derived from uncultivated virus populations observed within marine viromes. After magnetic bead purification, amplicon molecules were tagged with a unique barcode sequence using ligation. Purified, barcoded amplicons were then subjected to a limited cycle PCR with barcode sequences as template targets. Enriched barcoded amplicons from each sample were then purified, pooled in equal proportions, and sequenced using a PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA). Detailed methods describing DNA extraction, amplification of Class II RTPR target genes, and amplicon sequencing are provided in Supplementary Methods.

### DNA extraction from the bacterioplankton fraction, and 16S rRNA amplicon and tag-encoded Class II RTPR sequencing

DNA was extracted from the 0.22  $\mu\text{m}$  Sterivex filters (bacterioplankton fraction) using the phenol/chloroform method [58]. The V3–V4 hypervariable region of the 16S rRNA gene was PCR-amplified and sequenced from the bacterioplankton fraction on the Illumina MiSeq (Illumina, San Diego, CA) using a Nextera XT DNA Library Preparation Kit (Illumina), which exploits a dual-indexing strategy for multiplexed sequencing [59]. Class II RTPR genes within the bacterioplankton fraction were PCR-amplified, barcoded, and enriched using the same protocol as described for viroplankton. Detailed methods describing DNA extraction, amplification of 16S rRNA and Class II RTPR target genes, and amplicon sequencing are provided in Supplementary Methods.

### Class II RTPR amplicon quality control

General and detailed bioinformatics analysis pipelines are summarized in Fig. 1 and Supplementary Fig. S2, respectively. Briefly, Class II RTPR gene amplicon sequences from the viroplankton fraction (PacBio RSII sequencer) and the bacterioplankton fraction (PacBio Sequel sequencer) were initially screened for low quality bases and read length. Circular consensus sequence (CCS) reads were generated. Reads with less than three full passes, less than 98% minimum predicted accuracy, and with a length shorter than 250 bp or longer than 5,000 bp were excluded. Subsequently, sequence reads were demultiplexed, and primer and barcode sequences were detected and removed. Reads were translated into predicted amino acid sequences using a custom frameshift polishing pipeline ([https://github.com/dnasko/frameshift\\_polisher](https://github.com/dnasko/frameshift_polisher)). Key catalytic residues within translated amplicons were used for validating sequences as true Class II RTPRs using the Protein Active Site Validation (PASV) (<https://github.com/mooreryan/pasv>) pipeline [60]. Detailed methods describing bioinformatic processing of Class II RTPR amplicon sequences are provided in Supplementary Methods.

### Viroplankton Class II RTPR 98% OTUs and phylogenetic clades

Quality-controlled viroplankton Class II RTPR amino acid sequences were run through 11 *de novo* operational taxonomic unit (OTUs) clustering assessments varying the percent sequence identity (90–100% amino acid sequence identity) using the *cluster\_fast* command with default settings in USEARCH (version 11.0) [61]. This heuristic process identified 98% OTU clusters as a balance between optimal lumping and splitting of OTUs (Supplementary Fig. S3). Viroplankton Class II RTPR 98% OTU amino acid sequences were taxonomically divided into clades using a phylotyping method. OTUs were aligned using MAFFT (version 7.450) with the ginsi method. A phylogenetic tree of viroplankton Class II RTPR 98% OTUs was built using FastTree standard method, which employs a variant of the bootstrapping technique known as “approximately unbiased (AU) bootstrap” to assess the statistical support for the inferred tree topology (version 2.1.11) [62] and visualized using Iroki [63]. Viroplankton Class II RTPR phylogenetic clades were defined manually.

### 16S quality control and classification

General and detailed analysis pipelines are summarized in Fig. 1 and Supplementary Fig. S2. Demultiplexed, paired-end 16S rRNA reads from the bacterioplankton fraction were imported into QIIME2 (QIIME2 2019.1) [64] for quality filtering and generation of amplicon sequence variants (ASVs) using DADA2 [65]. Forward and reverse reads were truncated at positions 35 and 300 and at positions 35 and 295, respectively. 16S ASVs were taxonomically identified (Additional File 1) using the SILVA132 QIIME-compatible release (full length, seven-level taxonomy) [66] as a reference database in QIIME2. Sequences taxonomically assigned to Archaea (22 ASVs) and chloroplasts (230 ASVs) were removed prior to downstream analysis. Taxonomy was later confirmed using the SILVA 138 database using SILVA ACT (<https://www.arb-silva.de/aligner/>). One ASV, formerly designated as “Candidatus Actinomarina”, changed to “Cyanobacteria chloroplast” in the SILVA 138 database and was retained in subsequent analyses (Additional File 2).

For proportionality tests with bacterioplankton 16S ASVs and viroplankton Class II RTPR 98% OTUs, ASVs were restricted to only those that occurred in the 29 samples with viroplankton Class II RTPR amplicons. From that subset of 29 samples, only ASVs with a total count greater than 100 and with at least 10 counts in at least 10 samples were retained for downstream analysis. This avoided the risk of spurious associations based on rare populations observed in few samples.

### Class II RTPR and 16S community and population analyses

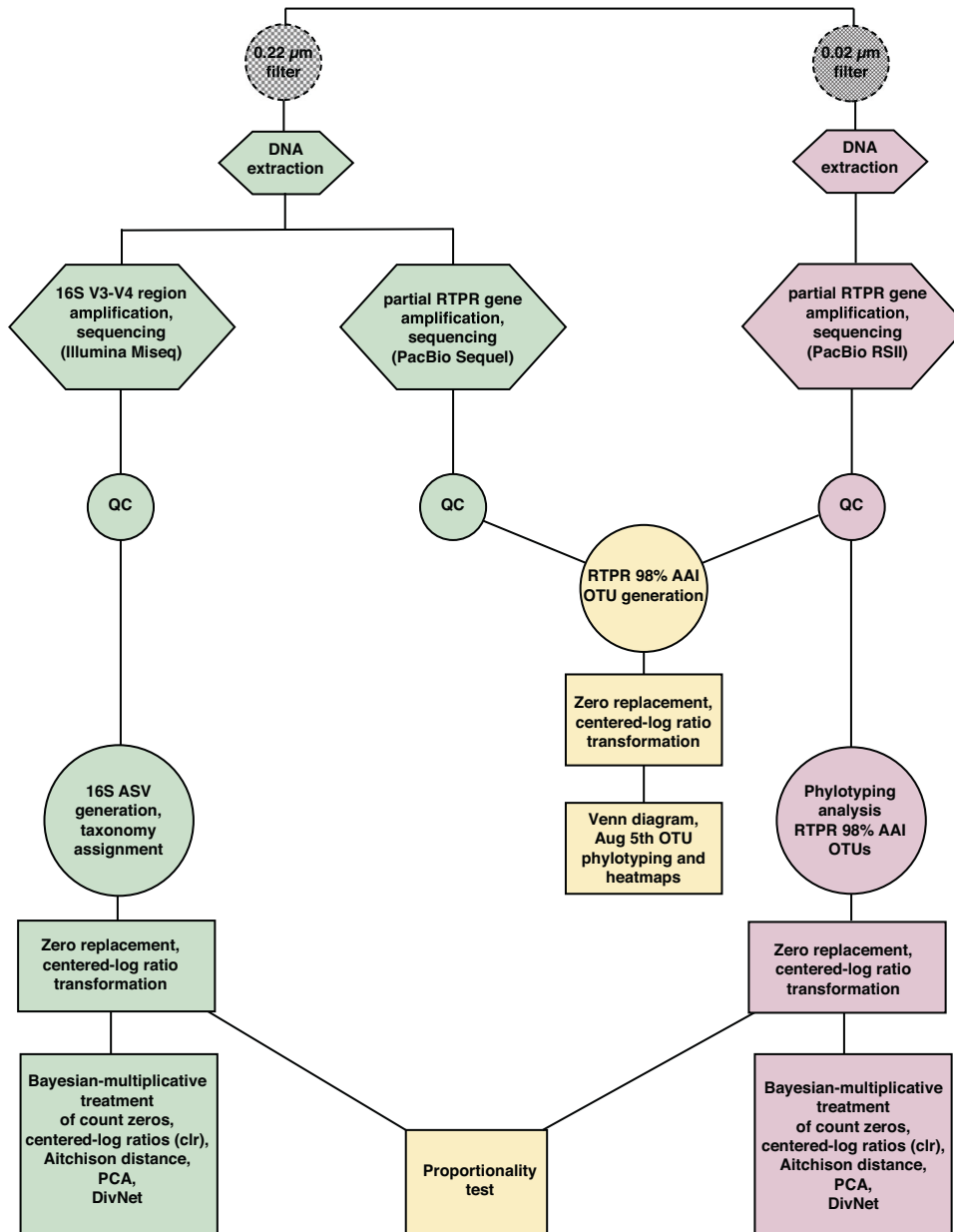
Viroplankton Class II RTPR 98% OTUs, viroplankton Class II RTPR phylogenetic clades, and bacterioplankton 16S ASVs were used in community and population analyses such as alpha and beta diversity, and population proportionality. The influence of environmental conditions on virio- and bacterioplankton communities were determined based on complete-linkage hierarchical clustering of Class II RTPR 98% OTUs, Class II RTPR phylogenetic clades, or 16S ASV community profiles using the hclust method in R. Tested environmental features include sampling day, depth, temperature, density, current velocity, salinity, nanoeukaryote abundance, picoeukaryote abundance, *Prochlorococcus* abundance, *Synechococcus* abundance, total autotroph abundance, viral abundance, and bacterial

abundance. Clustering of virio- and bacterioplankton communities was based on the Aitchison distance between the May 12th 5 m sample and every other sample. Resulting dendrograms were visualized with Iroki. Correlations between virio- or bacterioplankton communities and associated environmental features were tested using the *qiime diversity mantel* [67] command in QIIME2 (permutation times: 999). Significant correlations were defined as  $p < 0.05$ . Associated environmental features that showed significant correlations with sample community composition were plotted next to the sample community composition dendrogram.

Alpha diversity of the virio- or bacterioplankton communities in each sample was estimated based on the Shannon index using the R package DivNet (version 0.3.2) [68] and displayed using ggplot2 (version 3.3.0) [69]. Significance of alpha diversity differences was tested using beta function in the R package DivNet breakaway (version 4.6.14) [70]. Estimated Shannon indices were converted to the effective number of species (ENS, Hill numbers) [71]. Correlations of ENS estimates based on 98% OTUs or Class II RTPR phylogenetic clades were tested using the *cor.test* method in R. Beta diversity was visualized through principal component analysis

(PCA) plots of sample community compositions using the R package compositions (version 1.40.3) [72]. Significance of beta diversity differences between sampling day or depth groups was tested using *beta-group-significance* command (*permanova*) method [73] in QIIME2.

Associations within or between virioplankton Class II RTPR phylogenetic clades and bacterioplankton 16S ASV populations were explored using proportionality (*rho*,  $\rho$ ) tests, with  $-1 < \rho < 1$  [74]. Only positive  $\rho$  values reflecting a positive association between two populations were considered, as negative  $\rho$  values cannot be conclusively explained [75]. Only those associations with false discovery rates (FDRs) lower than 0.05 were retained for further analysis. Proportionality and FDR tests were performed using the R package *propr* (version 4.2.8). Associations with  $FDR < 0.05$  were extracted and plotted in a heatmap using ComplexHeatmap [76]. Columns and rows of the heatmap were clustered based on phylogenetic relations of virio- and bacterioplankton, respectively. Virioplankton Class II RTPR clades were arranged by their topological order on the 98% OTU phylogenetic tree. Bacterioplankton 16S ASVs were arranged by their 16S phylogenetic placement or association patterns.



**Fig. 1 Overview of experimental methods and analysis pipeline.** The 0.22 µm filter captured bacterioplankton cells. The 0.02 µm filter captured virioplankton particles. Hexagons represent molecular genetic experimental methods. QC stands for a collection of quality control steps specific to the type of next-generation sequencing technology used in sequencing amplicon libraries. Circles represent conventional data analysis methods. Squares represent compositional data analysis (CoDA) methods.



## Virio- and bacterioplankton population analysis using common Class II RTPR 98% OTUs

In 27 of 35 samples, Class II RTPR amplicons were obtained from both virio- and bacterioplankton fractions yielding 47,997 quality-controlled Class II RTPR amino acid sequences. Quality-controlled virio- and bacterioplankton Class II RTPR amino acid sequences were run through 11 *de novo* operational taxonomic unit (OTU) clustering assessments varying the percent sequence identity (90–100% amino acid sequence identity) using the *cluster\_fast* command with default settings in USEARCH. This heuristic process identified 5411 98% OTU clusters as a balance between optimal lumping and splitting of OTUs (Supplementary Fig. S4). OTUs with less than ten counts were removed prior to downstream analysis. The COUNTIF function in Excel was used for binning the remaining 254 Class II RTPR 98% OTUs into those containing sequences from both fractions, only the viroplankton fraction, or only the bacterioplankton fraction. Class II RTPR 98% OTU counts were calculated using the SUM function in Excel prior to zero replacement and clr transformation. An arbitrary clr value of 5 or –5 was assigned to those common 98% OTUs found exclusively in the virio- or bacterioplankton fraction, respectively. The difference in clr abundance between fractions of each Class II RTPR 98% OTU was calculated using the subtraction formula in Excel. Venn diagrams were plotted to show the total clr abundance of each OTU and the distribution of each OTU (difference in clr abundance) in the virio- and/or bacterioplankton fraction.

Only the August 5th sampling date yielded Class II RTPR amplicons from virio- and bacterioplankton fractions at all depths. Thus, this date was chosen for assessing depth-resolved distribution patterns of Class II RTPR viroplankton populations within the virio- and bacterioplankton fractions. Calculation of clr abundance values for the August 5th OTUs was performed by removing any OTU not present in at least one August 5th sample, leaving 1477 of the 5411 OTUs observed in the combined virio- and bacterioplankton datasets. Twenty-nine August 5th OTUs contained more than 50 sequences and had two or fewer zero replacements, and were retained for further analysis. Similarity between the phylogenetic distance of these 29 Class II RTPR 98% OTUs and their depth-resolved distribution patterns within the viroplankton fraction, the bacterioplankton fraction, and the combined virio- and bacterioplankton fraction were tested by plotting the cophenetic distance versus Aithison distance for all combinations of 29 OTUs. Combinations rather than permutations of OTUs were used, as the distances used are symmetric (e.g.,  $d(A, B) = d(B, A)$ ), where  $d$  is the distance function, and  $A$  and  $B$  are OTUs). Self-distances (e.g.,  $d(A, A)$ ,  $d(B, B)$ ) were not included. Linear models and significance tests were performed with the *lm* function in R.

## Data transformation

Compositional data analysis (CoDA) methods were applied to all analyses of community diversity and population interactions. Zero values in the feature tables [viroplankton Class II RTPR 98% OTUs (Additional File 3); Class II RTPR phylogenetic clade (Additional File 4); bacterioplankton 16S ASVs (Additional File 5); bacterioplankton 16S ASVs retained in the 29 samples with Class II RTPR amplicons (Additional File 6); and common Class II RTPR 98% OTUs found in both the bacterio- and viroplankton fractions (Additional Files 7 and 8)] were replaced by imputed values using the count zero multiplicative method from the R package *zCompositions* (version 1.3.3.1) [77] in R (version 3.5.3) [78]. The centered-log-ratio (clr) transform was applied to the zero-replaced dataset to produce clr abundance using the R package *CoDaSeq* (version 0.99.4) [79]. Within the August 5th subset (sampling date with Class II RTPR amplicons in both fractions at all depths) of common Class II RTPR 98% OTUs (Additional File 9), zero count values were replaced using *cmultRepl* with the method count zero multiplicative (CZM) in the *zCompositions* package (version 1.3.4) of R (version 4.0.2). Subsequently, the clr transformation was performed using base 2 in the R package *CoDaSeq*.

## Class II RTPR 98% OTUs within global RNR context

Class II RTPR 98% OTUs from both virio- and bacterioplankton fractions, Class II RTPRs identified through stringent homology search of the *Tara* Oceans dataset [80], and RNRdb [39] databases were aligned using Geneious 10.0.9 MAFFT alignment FFT-NS-2 and trimmed to the region of interest (H346 to S643 in *Lactobacillus leichmannii* monomeric class 2 adenosylcobalamin-dependent ribonucleotide-triphosphate reductase). A phylogenetic tree of trimmed sequences was built using FastTree and visualized with Iroki.

## RESULTS

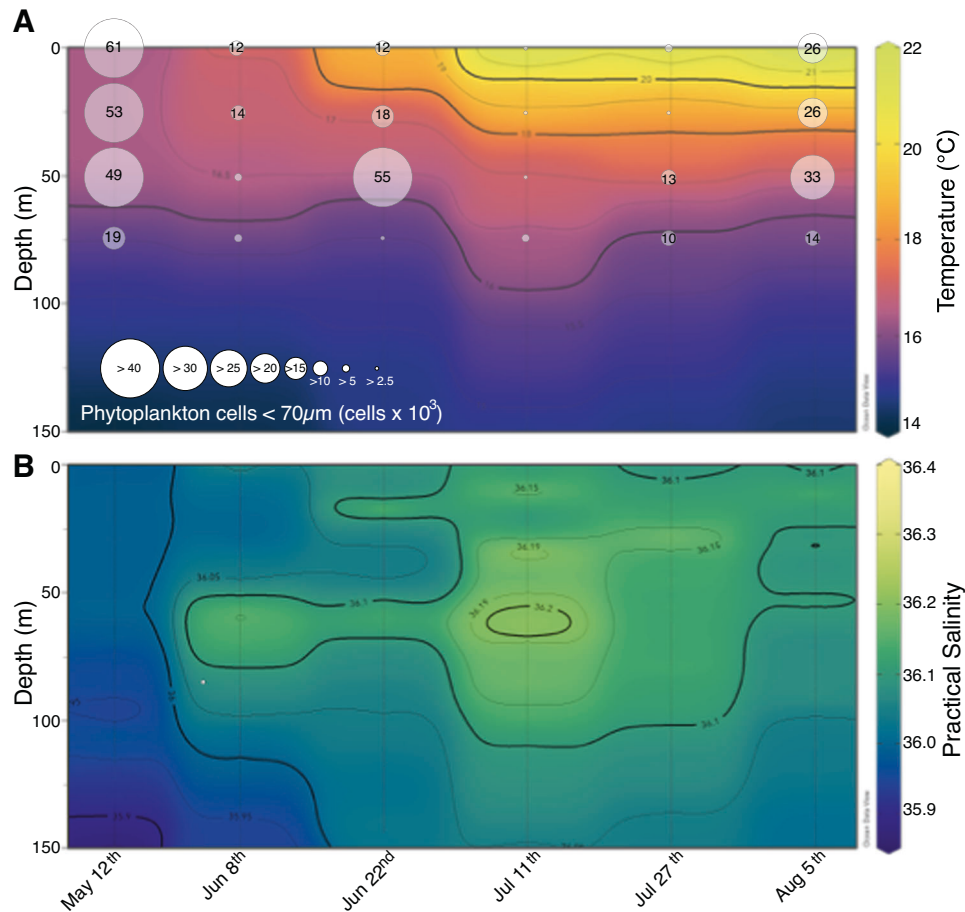
### Water column environment

Temperature of surface waters at station SCFP increased from May 12th (16.4 °C) to August 5th (22.0 °C) (Fig. 2A), agreeing with broader sea surface temperature (SST) data derived from satellite imagery (Supplementary Fig. S5A). Surface salinity (using the Practical Salinity Scale) increased from May 12th (35.9) to July 11th (36.1), decreased by 0.1 on July 27th, and increased again to 36.1 on August 5th (Fig. 2B). Surface waters were well mixed down to 50 m as indicated by temperature and salinity contours prior to June (Fig. 2). Water column stratification intensified throughout the sampling period, accompanying the mixed layer depth rising from 43 m (May 12th) to 5 m (August 5th). Salinity showed distinct sub-surface increases during the study period, indicating substantial water movement in the sampling region. Salinity changes in the sub-surface were centered at the 75 m depth, from June 8th to July 27th, especially on July 11th. The observations suggested the presence of an anticyclonic (downwelling) vortex at the sampling location around July 11th. Uplifts of isotherms and isohalines found in temperature and salinity profiles between sampling dates suggested the presence of cyclonic (upwelling) vortices, on June 22nd (more pronounced) and on July 27th (less pronounced) extending to August 5th. Satellite-derived data in 2016 (Supplementary Fig. S5A) showed one phytoplankton bloom, based on chlorophyll *a* and particulate organic carbon (POC), that occurred between February and April, with its maximum in March.

The sampling period started after the seasonal spring phytoplankton bloom and spanned the environmental and biological conditions observed from late spring to late summer. Observed changes in phytoplankton abundance corresponded with seasonal environmental changes during the sampling period (Fig. 2A and Supplementary Fig. S5B). High phytoplankton abundance ( $19\text{--}61 \times 10^3$  cells/mL) was observed throughout the water column from 0 to 75 m depth on May 12th, indicating the end of the annual spring bloom. Subsequently, phytoplankton abundance fell below  $2 \times 10^3$  cells/mL except for a sub-surface peak of  $\sim 55 \times 10^3$  cells/mL at 50 m on June 22nd, coinciding with the observed cyclonic upwelling event (Supplementary Table S1). The lowest phytoplankton abundance was observed on July 11th, coinciding with the anticyclonic downwelling event. Another increase in phytoplankton abundance coincided with the second, albeit less pronounced, cyclonic upwelling event during the subsequent two sampling dates, July 27th and August 5th.

### Changing environmental conditions and viroplankton and bacterioplankton community dynamics

For unknown reasons, only 29 of 35 viroplankton samples yielded Class II RTPR amplicons (Additional Files 3 and 10). From a total of 30,473 viroplankton Class II RTPR amino acid sequences, 3697 Class II RTPR 98% OTUs were identified (Supplementary Figs. S2 and S3). The Class II RTPR 98% OTUs clustered into 17 phylogenetic clades as manually determined by phylogenetic tree topology (Fig. 3). The three clades topologically located between clades 12 and 13 (gray color branches in Fig. 3) were excluded from downstream analyses due to their low abundance (7 OTUs, 12 amino acid sequences) and distant phylogenetic relations to other clades, leaving 3690 98% OTUs (30,461 amino acid sequences, Additional File 3) in 17 clades (Additional File 4). Phylogenetic clade 5 was the most abundant based on both amino acid sequence counts (3860) and clr abundance (1.07). The 17 Class II RTPR phylogenetic clades were used in subsequent compositional data analyses, including community diversity and population interactions analyses, as clades of related OTUs provided clearer ecological signals than the larger collection of more highly resolved Class II RTPR 98% OTUs. Class II RTPR 98% OTU community diversity within and between samples was also determined for comparison with the phylogenetic approach and these analyses are shown in supplemental figures.



**Fig. 2** Water column environment during the sampling period. Vertical profile contours of **A** temperature and phytoplankton cell numbers, and **B** salinity at station South of Channel Faial-Pico islands (SCFP) from May 12th to August 5th, 2016. Data was not collected on September 8th due to equipment malfunction. Vertical gray lines indicate the sampling date and the maximum depth of data collection. Temperature and salinity data between sampling dates was interpolated using the weighted average grid method. Circle sizes indicate phytoplankton abundances at each sampled depth (surface, 25, 50 and 75 m). Surface values were the average abundance at the 0 and 5 m depths.

The V3–V4 region of the 16S rRNA gene was PCR-amplified from 35 bacterioplankton fraction samples and sequenced (Additional Files 1, 2 and 10). After sequence quality filtering, 1779 amplicon sequence variants (ASVs) were identified from a total of 918,792 sequences across all samples (Additional File 5). Complete-linkage hierarchical clustering based on the Aitchison distance of clr-transformed sequence abundance between the first (May 12th 5 m) water sample and every other sample was performed for each of three datasets (virio- or bacterioplankton communities). Mantel tests assessed each Aitchison distance matrix with distance matrices of all collected environmental metadata (Supplementary Table S2) to identify environmental factors correlated with and potentially driving the clustering of virio- or bacterioplankton communities. Sample depth, collection day (displayed as Julian day), temperature, and water density were significantly (Spearman and Pearson  $p$  value  $< 0.02$ ) correlated with virioplankton community composition (Class II RTPR phylogenetic clades and Class II RTPR 98% OTUs) and bacterioplankton community composition. For the significant environmental variables, correlation coefficients with virioplankton populations were generally higher when using the Class II RTPR phylogenetic clades.

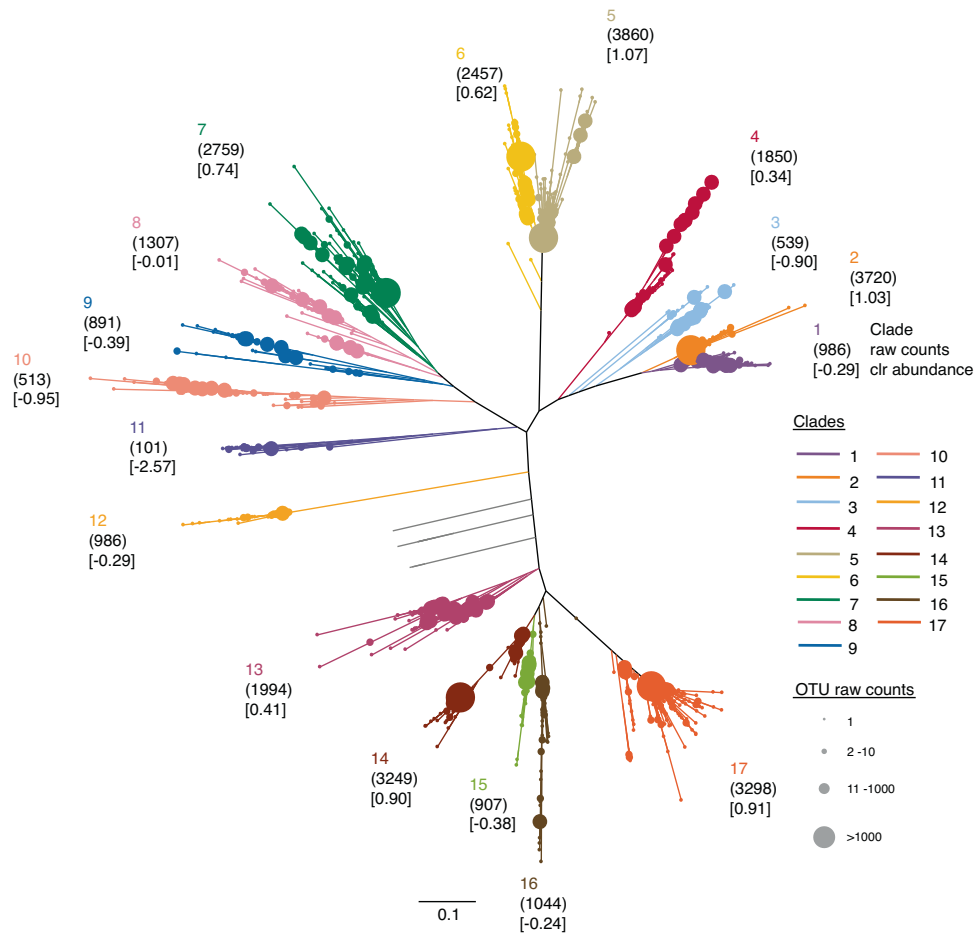
Hierarchical clustering of virioplankton communities based on Class II RTPR phylogenetic clades formed four large groups (vA, vB, vC, and vD) (Fig. 4A) based on their distance from the May 12th 5 m sample. Community groups vA/vB and groups vC/vD were organized into equally distant supergroups. Within each group,

communities had similar profiles for depth/density and for temperature/day. Groups vA/vB contained communities generally occurring at deeper depths with lower water temperatures, whereas groups vC/vD contained communities generally collected at shallower depths with higher water temperatures. However, for all groups, there were outliers likely reflecting some of the low rank correlation coefficients (0.19–0.51) between community structure and environmental metadata (Supplementary Table S2). Bacterioplankton communities formed three groups (bA, bB, and bC) (Fig. 4B) based on their distance to the May 12th 5 m sample. Communities in groups bA and bB were observed at lower water temperatures, whereas group bC communities were found in higher water temperatures at shallow depths ( $< 25$  m). The division between groups bA and bB was more subtle and may have been forced by depth as group bB contained communities in only 50 and 75 m samples.

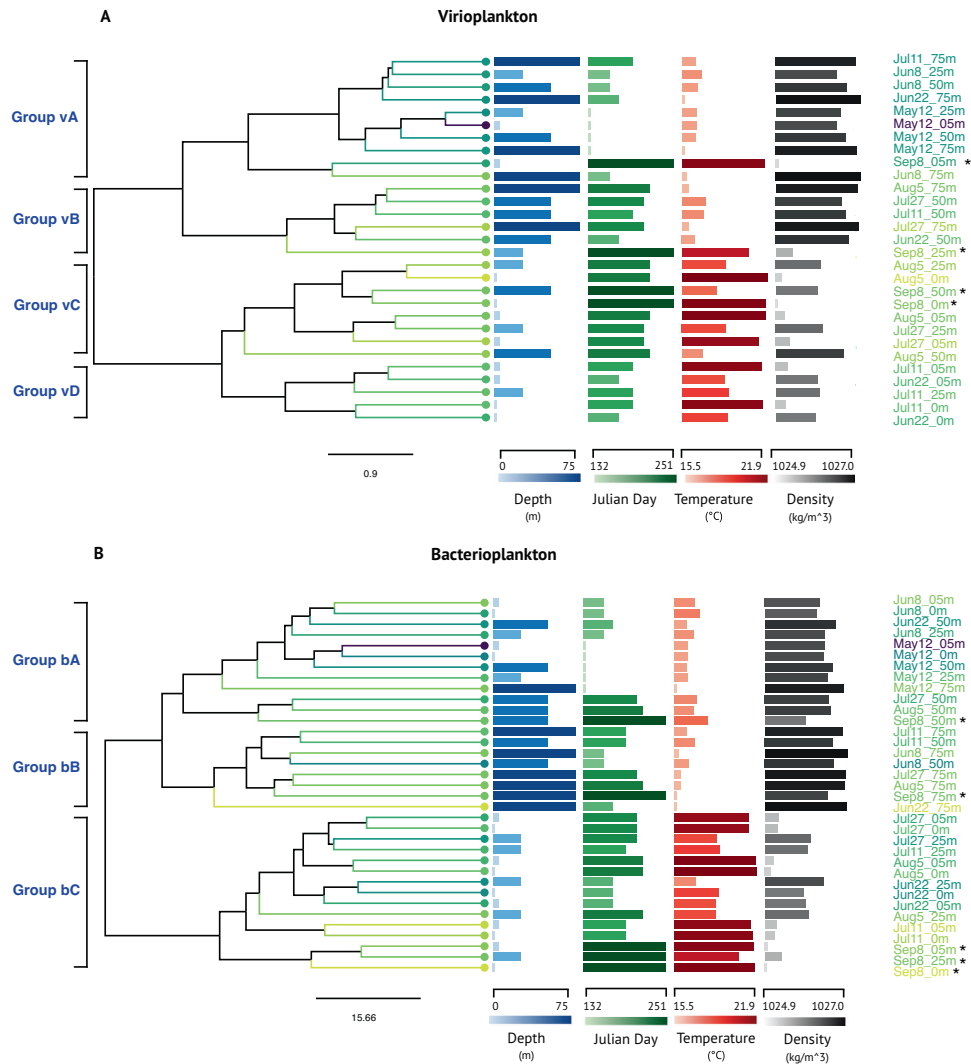
The effects of sampling date (Fig. 5A) and depth (Supplementary Fig. S6A) on viral and bacterial ENS were significant (beta  $< 0.05$ ; data not shown); therefore, these environmental parameters were selected for analysis of their interactions with alpha and beta diversity of the virio- and bacterioplankton communities. Over the interseasonal sampling period, virioplankton alpha diversity according to Class II RTPR phylogenetic clades (Fig. 5A) fluctuated slightly reaching a maximum around the July 11th anticyclonic downwelling event (Fig. 2). This peak was shouldered by minimum diversity estimates on June 8th and July 27th corresponding with cyclonic upwelling events. Virioplankton alpha

diversity fluctuated slightly with depth (Supplementary Fig. S6A). Maximal and minimal diversity estimates occurred at 25 and 75 m, respectively. ENS estimates of alpha diversity based on 98% OTUs (Supplementary Fig. S7A and B) were three to six-fold higher than ENS estimates for virioplankton based on Class II RTPR phylogenetic clades (Fig. 5A and Supplementary Fig. S6A) as there were thousands of 98% OTUs versus only 17 Class II RTPR phylogenetic clades. The general trends of ENS estimates based on 98% OTUs and Class II RTPR phylogenetic clades were consistent (Spearman's rank correlation:  $\rho = 0.82$ ,  $p$  value = 0.034; Pearson's product-moment correlation:  $r = 0.78$ ,  $p$  value = 0.037). Increases in Class II RTPR 98% OTU ENS (Supplementary Fig. S7A) were observed on June 22nd and July 11th, corresponding with the observed cyclonic upwelling event. The maximum ENS was observed on September 8th; unfortunately its association with an upwelling event is unknown as equipment failure prevented temperature measurements on this date. The maximum and minimum virioplankton Class II RTPR 98% OTU ENS in samples by depth were 5 and 50 m, respectively (Supplementary Fig. S7B). Bacterioplankton alpha diversity steadily declined from its May 12th peak to a minimum value on July 11th followed by a slight increase over the remaining sampling dates (Fig. 5A). In contrast to virioplankton, bacterioplankton alpha diversity did change dramatically with depth demonstrating steady increases from minimal diversity estimates in surface waters to the maximal diversity estimates observed at 75 m (Supplementary Fig. S6A).

Relationships between virioplankton Class II RTPR phylogenetic clade communities were examined using principal components analysis (Fig. 5B). Sixty-five percent of the variability between communities was explained by the first two principal components (PC1 48.9%, PC2 16.0%), and sampling date (Fig. 5B) seemed to explain this variability more than sample depth (Supplementary Fig. S6B). Using the 98% OTU populations, ~35% of the variability between communities was explained by the first two principal components (PC1 22.8%, PC2 12.1%), and was associated with both sampling date and depth (Supplementary Fig. S7C, D). The variability in virioplankton communities explained by the Class II RTPR 98% OTUs was approximately half of that explained by phylogenetic clades. This trend was attributed to the higher number and granularity of Class II RTPR 98% OTU populations in describing virioplankton community structure. For either means of defining virioplankton communities, PCA analyses showed that the May 12th virioplankton communities grouped closer to the origins of both the PC1 and PC2 axes, apart from most other sampling dates, and the August 5th samples tended to place toward the ends of PC1 and PC2 axes. Virioplankton communities as defined by Class II RTPR phylogenetic clades demonstrated significant differences based on date and depth (PERMANOVA  $p$  value < 0.05; Supplementary Table S3A, D), largely agreeing with observations of 98% OTUs (Supplementary Table S3B & E). May 12th virioplankton communities differed from most other time-points. June 8th communities differed from July 27th, and August



**Fig. 3** Approximate maximum likelihood tree of virioplankton Class II RTPR 98% OTU representative sequences. Shared branch, node, and clade number colors indicate Class II RTPR phylogenetic clade manual assignment. Seven OTUs (between clades 12 and 13) with low abundance and distant relationships with other phylogenetic clades are represented by gray branches and were excluded from downstream analysis. Node sizes indicate 98% OTU counts. Each phylogenetic clade's total amino acid sequence counts and CLR abundance are indicated in rounded and squared parentheses, respectively. Scale bar indicates the average number of amino acid substitutions per site.



**Fig. 4** The effects of environmental factors on viroplankton and bacterioplankton community structures. Phylogenetic clustering of viro- and bacterioplankton communities (dendrograms) based on **A** viroplankton Class II RTPR phylogenetic clades and **B** bacterioplankton 16S ASVs, alongside environmental variables (columns of horizontal bars) demonstrating significant correlations (Spearman and Pearson tests,  $p < 0.05$ ). Dendrograms are based on complete-linkage hierarchical clustering of each sample's Aitchison distance to the sample collected May 12th at 5 m (May12\_05m) based on centered-log-ratio (clr) abundance. Scale bar indicates Aitchison distance between samples. Major groups of viro- or bacterioplankton communities are labeled to the left of each dendrogram. Environmental variables are represented in columns, with horizontal bars indicating the measurement for each sample both by color gradient and by bar length compared to the numerical scale along each column's horizontal axis [Depth (blue), Day (green), Temperature (red), and Density (gray)]. Sample labels with asterisks (\*) indicate samples where temperature and density were estimated using the Marine Copernicus Global Ocean Physics model.

5th communities differed from those on June 8th, June 22nd, and July 11th. Depth-based viroplankton community structure (Supplementary Fig. S6B) did not group as strongly on the PCA plots as date-based communities (Fig. 5B). Nevertheless, surface communities (0 m) were different from those at 50 and 75 m, and 5 m communities differed from 75 m communities.

Relationships between bacterioplankton 16S ASV communities were examined using principal components analysis as well (Fig. 5C). Approximately 38% of the variability between communities was explained by the first two principal components (PC1 25.2%, PC2 12.4%), and was associated with both sampling date (Fig. 5C) and depth (Supplementary Fig. S6C). As seen with viral communities defined by either Class II RTPR phylogenetic clades or 98% OTUs, the May 12th bacterioplankton samples grouped toward the origin of both PC1 and PC2, apart from most other sampling dates. The August 5th bacterioplankton communities showed a greater spread across PC1 than the viroplankton communities for this sampling date. These observations were

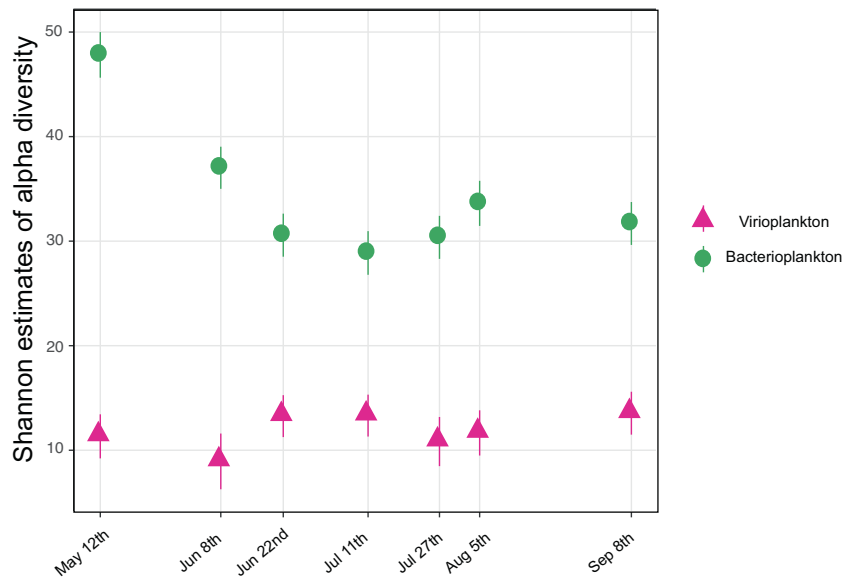
statistically confirmed. Bacterioplankton communities on May 12th and June 8th were different from those observed on all other sampling dates, except the pair of samples from June 8th and June 22nd (PERMANOVA  $p$  value and/or  $q$  value  $< 0.05$ ; Supplementary Table S3C). Bacterial 16S ASV communities were also separated by depth, with 50 m and 75 m communities occurring along the left side of PC1 (Supplementary Fig. S6C). Surface bacterioplankton communities at 0 and 5 m significantly diverged from the deep communities at 50 and 75 m; 25 m communities diverged from the 75 m communities; and 50 m communities diverged from the 75 m communities (PERMANOVA  $p$  value and/or  $q$  value  $< 0.05$ ; Supplementary Table S3F).

#### Associations between viroplankton Class II RTPR phylogenetic clades and bacterioplankton 16S ASVs

The proportionality test was used for comparing changes in the clr abundance of 17 viroplankton Class II RTPR populations (as defined by phylogenetic clades on Fig. 3) and the subset of 221

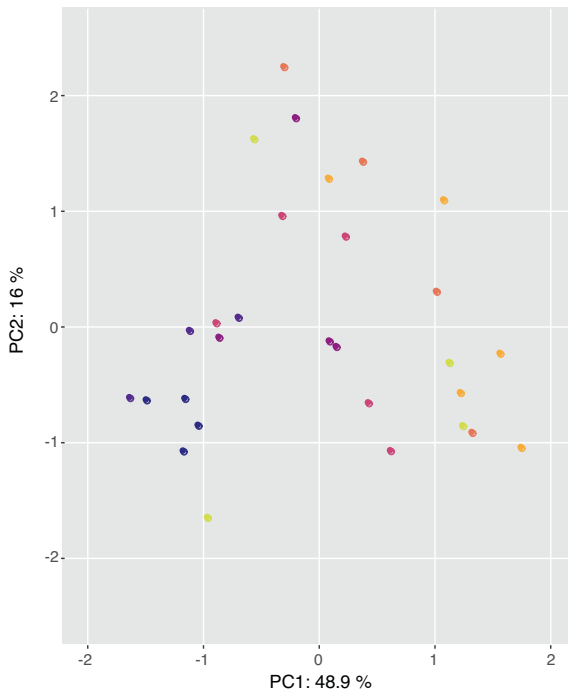


A



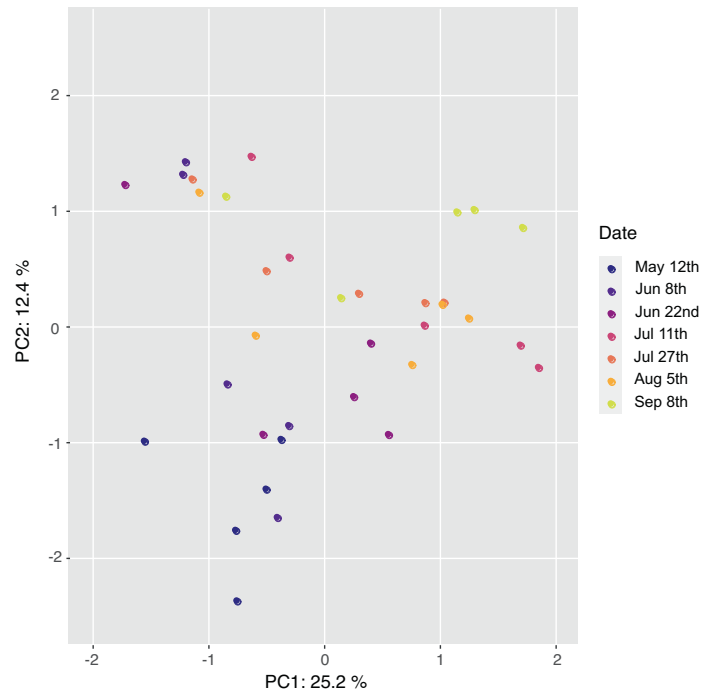
B

Virioplankton clade



C

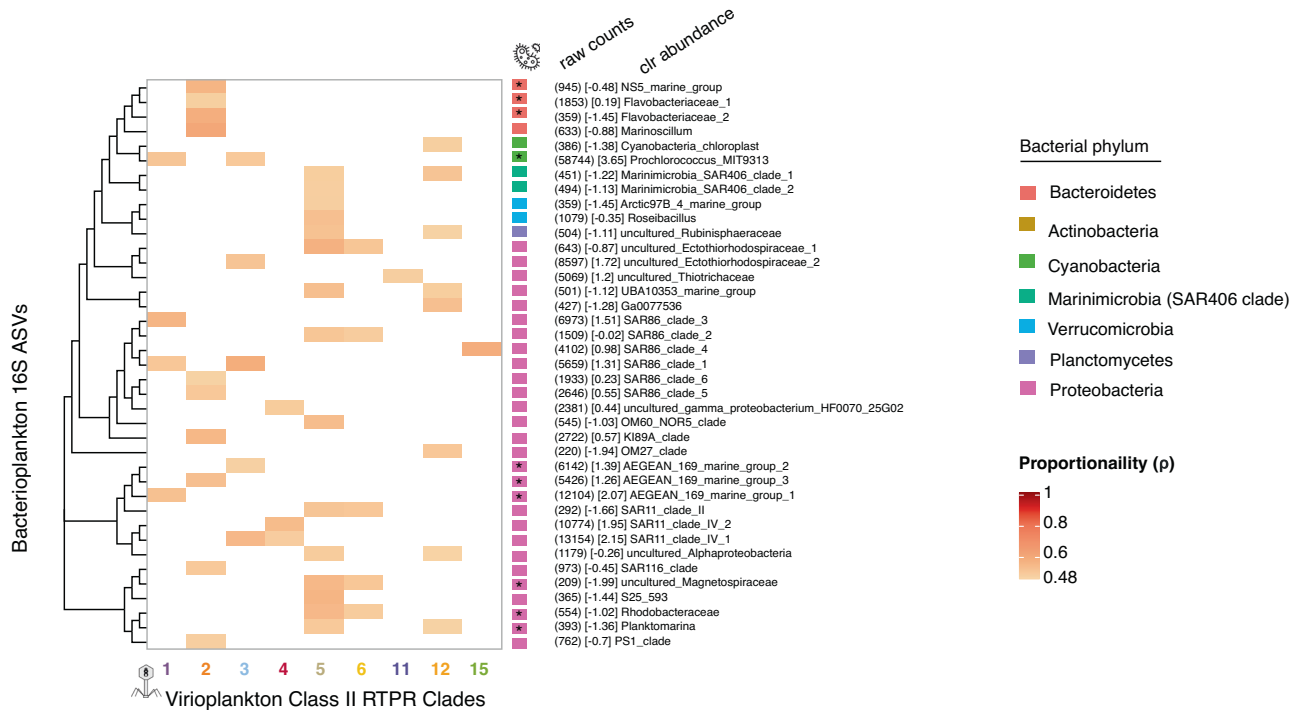
Bacterioplankton ASV



**Fig. 5 Sampling date influenced virioplankton (Class II RTPR phylogenetic clade) and bacterioplankton (16S ASV) community alpha and beta diversity.** **A** Effective number of species (ENS, based on Shannon indices, a measurement of alpha diversity) within virioplankton Class II RTPR and bacterioplankton 16S communities by sampling date. Error bars represent two standard deviations of the estimates. Principal component analysis (PCA) plots of beta diversity based on centered-log-ratio (clr) transformed abundance of **B** virioplankton RTPR phylogenetic clade communities or **C** bacterioplankton 16S ASV communities. Samples (circles) are colored according to sampling date.

bacterioplankton 16S ASV populations in 27 samples with both virioplankton Class II RTPR and bacterioplankton 16S amplicons (Additional File 11). Virio- and bacterioplankton population pairs with  $\rho$  values  $\geq 0.48$  had positive associations as the false discovery rates were  $\leq 0.05$  (Supplementary Fig. S8). Nine of the

17 virioplankton Class II RTPR phylogenetic clades had positive associations ( $\rho \geq 0.48$ ) with 39 of 221 bacterioplankton ASV populations (Fig. 6). Associating virioplankton clades were phylogenetically diverse and were observed at both high (clade 2 and 5; clr 1.03 and 1.07, respectively) and low abundance (clade



**Fig. 6 Matrix (p proportionality) of significantly associated ( $p \geq 0.48$ ) virioplankton Class II RTPR phylogenetic clades (x-axis) and bacterioplankton 16S ASVs (y-axis) (Additional File 11).** Virioplankton Class II RTPR clades are colored and arranged by their topological order on the 98% OTU phylogenetic tree (Fig. 3). Bacterioplankton 16S ASVs are arranged according to phylogeny. Bacterioplankton 16S ASV phylum (colored block), raw sequence count, clr abundance, and taxonomic assignment are shown on the right. Asterisk indicates those ASVs within orders, families, or genera containing species capable of  $B_{12}$  synthesis according to a pangenomic survey by Heal et al. [92].

3 and 11; clr  $-0.90$  and  $-2.57$ , respectively) (Fig. 3). Virio- and bacterioplankton abundance did not appear to drive these associations as low-abundance Class II RTPR virioplankton associated with high-abundance bacterioplankton and vice versa (Supplementary Fig. S9). The phylogenetic diversity of positively associating bacterioplankton populations was broad, encompassing seven bacterial phyla observed among the included 16S ASV populations. The bacterioplankton 16S ASV populations ranged in clr abundance from the most abundant taxon 3.65 (Prochlorococcus\_MIT9313) to one of the least abundant taxon  $-1.99$  (SAR116\_clade).

Virioplankton clade 5, the most abundant Class II RTPR phylogenetic clade (clr 1.07, Fig. 3), showed the most positive associations across a wide phylogenetic breadth of bacterioplankton, encompassing four of nine bacterial phyla and 15 of the 39 total positively associated 16S ASVs, including SAR11\_clade\_II (Fig. 6). Virioplankton Class II RTPR phylogenetic clades 1 and 3 associated with the most abundant bacterioplankton 16S ASV, Prochlorococcus\_MIT9313, and several 16S ASVs from Proteobacteria, such as SAR86, SAR11, and the Aegan169 marine group, well known to be abundant within oceanic ecosystems (Fig. 6). Virioplankton clade 2 showed exclusive associations with multiple bacterioplankton ASVs from Proteobacteria and Bacteroidetes despite the close phylogenetic associations of clades 1, 2, and 3 (Fig. 3). In contrast to virioplankton clades associating with multiple ASVs, clades 11 and 15 each associated with only a single bacterioplankton 16S ASV, uncultured\_Thiotrichaceae and SAR86\_clade\_4, respectively (Fig. 6).

Fourteen out of 136 non-self-comparisons between virioplankton Class II RTPR populations showed a positive association (Supplementary Fig. S10, Additional File 11). These associations between virioplankton clades did not follow a pattern according to phylogenetic distance as only three of the 14 associations were between neighboring Class II RTPR clades (clades 2 and 3; clades 5

and 6; and clades 8 and 9). Of the 221 bacterioplankton 16S ASV populations, 36 showed positive, non-self associations with other ASV populations (Supplementary Fig. S11, Additional File 11). One quarter (9 of 36) of these populations belonged to taxonomic groups containing known  $B_{12}$  producing species, the essential co-factor for the Class II RTPR protein. Overall, there were 390 positive non-self associations among the 36 ASVs,  $\sim 5\%$  of all the non-self comparisons (1260). Bacterioplankton populations were clearly divided by their association with either SAR11\_clade\_II or Prochlorococcus\_MIT9313 (red boxes in Supplementary Fig. S11). No 16S ASVs associated with both SAR11\_clade\_II and Prochlorococcus\_MIT9313. SAR11\_clade\_II associated with only other low-abundance (clr < 0) ASVs, whereas Prochlorococcus\_MIT9313 was associated with mostly high-abundance (clr > 0) ASVs. Beyond the delineation of SAR11\_clade\_II from Prochlorococcus\_MIT9313, positively associating bacterioplankton were divided into three larger groups (Supplementary Fig. S11 x-axis), defined by either associations with SAR11, or with Prochlorococcus, or a small third group consisting of only three Proteobacterial taxa showing a low number of associations without a consistent pattern.

#### Depth- and time-resolved viral-host associations

Spatiotemporal patterns in bacterioplankton associations with clade 5, the most abundant Class II RTPR virioplankton population, were explored by plotting changes in clr abundance at each sampling depth over time (Supplementary Fig. S12). Clade 5 Class II RTPR viral populations showed their highest abundances at sampling depths of 25 m and deeper, with consistently high abundances occurring at 75 m. Deeper samples also contained a greater number and diversity of bacterioplankton ASV taxa showing associations with clade 5 Class II RTPR virioplankton. Among the sampling dates, July 11th was noteworthy as several of the ASV taxa associated with clade 5 Class II RTPR virioplankton showed dramatic reductions in clr abundance corresponding with

the anticyclonic downwelling event. Correspondingly, the clr abundance of clade 5 viroplankton held steady or declined on the two subsequent sampling dates (July 27th and August 5th). A similar spatiotemporal analysis of associations between viro- and bacterioplankton populations was conducted for abundant bacterioplankton populations, *Prochlorococcus\_MIT9313* and three SAR11 clades, and four associating Class II RTPR viroplankton populations (Supplementary Fig. S13). Clades 1 and 3 Class II RTPR viroplankton populations showed associations with *Prochlorococcus\_MIT9313* (Fig. 6). Within a sampled depth across time, the clr abundance of clade 1 and 3 viruses was relatively steady, however, clr abundance of these two populations decreased in the 75 m samples as did the abundance of their associated *Prochlorococcus\_MIT9313* population. Class II RTPR clade 3 was also associated with SAR11\_clade\_IV\_1 which, like *Prochlorococcus\_MIT9313*, was most abundant in surface waters. Class II RTPR clade 4 showed positive associations with SAR11\_clade\_IV\_1 and \_2 (Fig. 6), which were the second and third most abundant bacterioplankton taxa showing associations with Class II RTPR viroplankton. While these SAR11 populations were most abundant at the surface (clr values of around 3.75), they were also abundant throughout the water column and changed only modestly with time. The behavior of Class II RTPR clade 4 showed a similar pattern as it was most abundant in surface waters (clr values  $\geq 1$ ) and then lower abundance at 50 and 75 m (clr values  $\sim 0$ ). Class II RTPR clade 5 associated with SAR11\_clade\_II which was the rarest of the SAR11 taxa showing associations with a Class II RTPR viroplankton population. SAR11\_clade\_II populations increased with depth, showing their highest abundances (clr values  $\sim 0$ ) at 75 m. Similarly, Class II RTPR clade 5 was less abundant at shallow depths (clr values  $< 1$ ) and then increased to its highest abundances at 75 m (clr values  $> 1$ ).

#### Free and actively replicating viral populations: Class II RTPR 98% OTUs occurring in viro- and bacterioplankton fractions

Twenty-nine viroplankton and 31 bacterioplankton samples provided Class II RTPR amplicons with an overlap of 27 samples providing amplicons from both. These 27 paired samples provided 47,997 Class II RTPR quality-filtered amino acid sequences, which were clustered *de novo* into 5411 Class II RTPR common 98% OTUs (Additional File 7). After removing 98% Class II RTPR OTUs with less than ten sequences, 21 OTUs were observed only in the viroplankton fraction, 40 only in the bacterioplankton fraction, and 193 in both fractions (Fig. 7 and Additional File 8). An encompassing phylogenetic analysis of these Class II RTPR 98% OTUs, sequences from the RNR database, and viromes from the *Tara* Oceans expedition showed that the Class II RTPR 98% OTUs from the viro- and bacterioplankton fractions occurred within a section of the tree containing few sequences from known bacteria or viruses (RNRdb) but several sequences from *Tara* Oceans viromes (Supplementary Fig. S14). The most abundant OTUs were primarily found in the shared fraction (Fig. 7A). Less abundant OTUs were largely found in either viro- or bacterioplankton fractions only. Examining the relative contribution of viro- or bacterioplankton amplicon sequences to Class II RTPR 98% OTUs required assignment of an arbitrary clr of 5 and  $-5$  to those OTUs found only within the viro- or bacterioplankton fractions, respectively. The difference in clr abundance within each Class II RTPR 98% OTU (clr(viroplankton)-clr(bacterioplankton), Fig. 7B) found that within most of the shared OTUs, viroplankton contributed a greater number of amplicons (66% of all shared OTUs (128 of 193), red colored OTUs above the dashed line in Fig. 7B).

All samples (fractions and depths) from the August 5th sampling date provided Class II RTPR amplicon sequences, enabling coordinate examination of viroplankton populations observed both as free viruses (i.e., within the viroplankton fraction) and viruses replicating within cells (i.e., within the

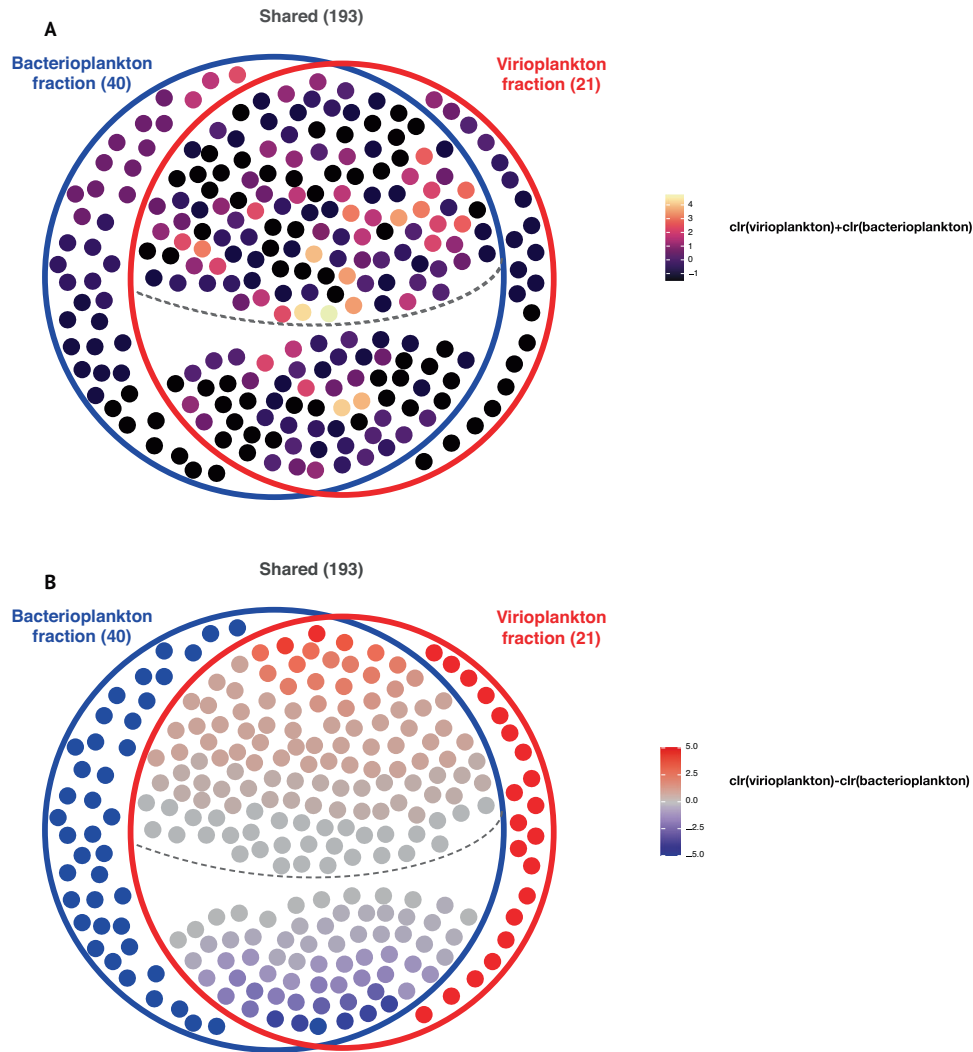
bacterioplankton fraction). Twenty-nine Class II RTPR 98% OTU populations were selected, each containing greater than 50 total sequences with two or fewer zero-replaced samples (Additional File 9). These heuristics prevented the inclusion of rare and inconsistently observed viroplankton populations. Phylogenetic analysis of these 29 Class II RTPR 98% OTUs alongside observations of clr abundance of each OTU within the viro- and bacterioplankton fractions addressed the hypothesis that the evolutionary distance between viroplankton populations reflects similarity in their ecological behavior (Fig. 8). Statistical testing comparing the cophenetic distance of the OTUs with the Aitchison distance of the OTUs according to their clr abundances across the August 5th samples demonstrated a significant ( $p = 1.6 \times 10^{-5}$ ), weak ( $R^2 = 0.045$ ), positive correlation (correlation  $r = 0.212$ ) between evolutionary distance and clr abundance in the bacterioplankton fraction (Supplementary Fig. S15). There was no significant relationship between cophenetic distance and clr abundance in the viroplankton fraction or when combining the viro- and bacterioplankton fractions (data not shown). Thus, more closely related Class II RTPR viroplankton populations demonstrated similar ecological behavior in terms of abundance as viruses replicating within cells (i.e., the bacterioplankton fraction), but not as free viruses.

The 29 August 5th Class II RTPR 98% OTUs occurred across nearly all of the major Class II RTPR clades identified in Fig. 3, with the exception of clades 10 and 11 which were among the rarest clades observed in the study (clr values  $-0.95$  and  $-2.57$ , respectively). Clade 7 was the most represented among the August 5th OTUs. These OTUs demonstrated a distinctive evolutionary split with a group of closely related OTUs (9, 15, 7, 16 and 26) clustering away from OTU 25 (Fig. 8). The clr abundance heat maps of the closely related group were similar with the majority of sequences occurring in the viroplankton fraction above 25 meters. In contrast, OTU 25 sequences were more abundant below 25 m demonstrating high abundance in the bacterioplankton fraction at 50 m. By and large, closely related August 5th OTUs showed similar clr abundance patterns supporting the finding of a significant relationship between evolutionary distance of Class II RTPR viroplankton populations and Aitchison distance.

## DISCUSSION

### Ribonucleotide-triphosphate reductase: a hallmark gene within viruses

As the only enzyme capable of reducing ribonucleotides to deoxyribonucleotides, RNR effectively controls the rate of DNA synthesis [33]. Thus, it makes intuitive sense that RNR genes would be common within the genomes of lytic viruses as speed and efficiency in genome replication is under positive selection pressure for viruses exhibiting a virulent lifecycle. Indeed, RNR genes are among the most commonly observed genes within dsDNA viral genomes and viromes, lending support to this hypothesis [9, 34, 81, 82]. RNRs are ancient proteins as enzymatic ribonucleotide reduction was a critical step from the RNA world to the emergent dominance of DNA as the principal information-carrying molecule for life on Earth [83–85]. Thus, it is possible that RNRs existed within viral genomes not long after the emergence of DNA. Phylogenetic analysis of known and environmental Class II RTPRs (Supplementary Fig. S14) indicated that viral Class II RTPR proteins diverged from cellular Class II RTPRs and are as deeply branching, reflecting both the unique selective pressure on viral Class II RTPRs and their long evolutionary history within viruses. However, the tree also shows the emergence of cellular Class II RTPRs from clades dominated by viral sequences and vice versa. These instances may represent viral to cellular (and cellular to viral) gene transfer events. Throughout the long evolutionary history of RNRs there is evidence of gene transfer between cellular



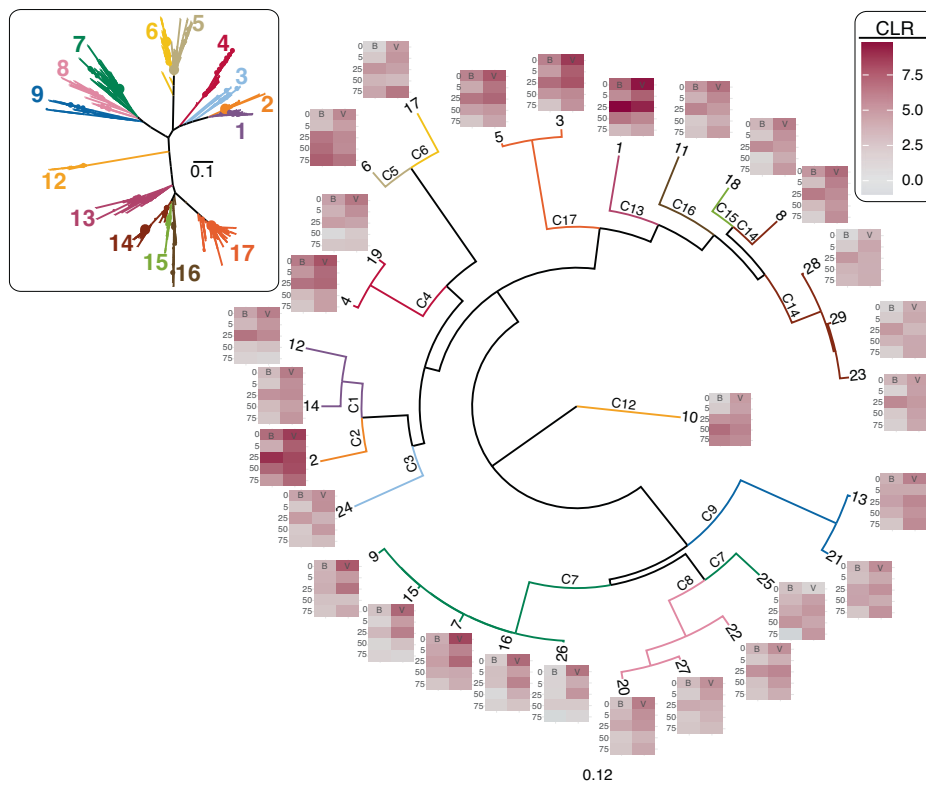
**Fig. 7 Class II RTPR 98% OTUs ( $n = 254$ ) identified from amino acid sequences present in 27 paired samples with amplicons from the virio- and bacterioplankton fractions.** Class II RTPR 98% OTUs (filled circles) were observed in either only the bacterioplankton fraction (blue circle,  $n = 40$ ), only the virioplankton fraction (red circle,  $n = 21$ ), or both (Shared,  $n = 193$ ). The dashed line in the overlap region divides the shared OTUs more abundant in the virioplankton fraction (above the line,  $n = 128$ ) from those more abundant in the bacterioplankton fraction (below the line,  $n = 65$ ). **A** Class II RTPR 98% OTUs are colored by their total centered-log-ratio (clr) abundance in both fractions ( $\text{clr}[\text{virioplankton}] + \text{clr}[\text{bacterioplankton}]$ ). **B** Class II RTPR 98% OTUs are colored by the difference between clr abundance in the virioplankton and bacterioplankton fractions ( $\text{clr}[\text{virioplankton}] - \text{clr}[\text{bacterioplankton}]$ ). Red colors indicate greater abundance of that OTU in the virioplankton fraction (values  $> 0$ , where 5 is an arbitrary maximum representing OTU presence in the virioplankton only). Blue colors indicate greater abundance of that OTU in the bacterioplankton fraction (values  $< 0$ , where  $-5$  is an arbitrary minimum representing OTU presence in the bacterioplankton only).

kingdoms, most notably the transfer of Class I RNRs from bacteria to archaea and ultimately eukaryotes [31] where Class I RNRs predominate (Supplementary Fig. S1A). In the case of nucleotide metabolism proteins, such as RNRs, it is possible that greater levels of biochemical innovation occur within lytic viruses than within cells, as viruses typically carry one allele of a particular nucleotide metabolism gene which is under intense selective pressure for fast and efficient DNA replication. Certainly, at the level of RNR class, selective processes may influence the distribution of RNR classes across the tree of life, where viruses demonstrate a distinctly high frequency of Class II RTPRs in comparison with cellular kingdoms.

Interestingly, a breadth of microbial diversity was associated with several of the Class II RTPR viral phylogenetic clades (Fig. 6). There are a few possible explanations for this. One possibility is that physiochemical factors were stronger drivers of microbial dynamics than viral-induced mortality, resulting in positive associations between Class II RTPR viral clades and a sub-

community of similarly adapted organisms, both host and non-host alike. However, it is also possible that the diverse taxa positively associated with several of the Class II RTPR viral clades reflect the hosts for viruses within these clades. Abundant marine phages, like 37-F6, have been predicted to infect hosts spanning different phyla [86], while direct linkages between viral RNR genes and 16S rRNA genes from diverse bacterial taxa have been observed [87]. In the latter case, viruses associated with diverse taxa were less efficient at infecting these hosts than viruses with narrow host ranges [87], but even infrequent cross-taxa infections could provide ample opportunities for horizontal gene transfer (HGT) of RNR genes between viral populations, resulting in similar RNR sequences across a diverse group of viruses infecting different hosts. Furthermore, horizontal gene transfer of RNRs across all three domains of life [31], as well as between phage populations [34], is well-established, suggesting that HGT has been critical to shaping viral ecology and cobalamin cycling in





**Fig. 8** Phylogram of 29 Class II RTPR 98% OTUs occurring within both the bacterioplankton (B) and viroplankton (V) fractions on the August 5th sampling date, with heat maps at nodes representing OTU centered-log-ratio (clr) abundance in each sample (fraction and depth). Node labels indicate OTU number. Branches are colored and labeled according to membership of the OTUs in the Class II RTPR clades (C1–C9; C12–C17) identified in the inset tree (Fig. 3). Scale bar indicates the average number of amino acid substitutions per site.

marine environments. Answering questions surrounding cellular–viral gene exchanges and their contribution towards the emergence of biochemical innovation will improve understanding of the evolutionary mechanisms behind adaptation in microbial communities.

### Biochemistry constrains Class II RTPR viroplankton population ecology

In the case of Class II RTPR, this monomeric Class II RNR requires an adenosylcobalamin ( $B_{12}$ ) co-factor (like all Class II RNRs), can perform ribonucleotide reduction under aerobic or anaerobic conditions, and, along with Class III RNRs, utilizes a ribonucleotide-triphosphate (NTP) substrate [88]. The  $B_{12}$  requirement and use of NTP substrates have intriguing implications for the ecology of viruses carrying Class II RTPR. Since cobamide nutrient (including  $B_{12}$ ) synthesis is complex, specialized, and energy intensive, few marine bacterial and archaeal phyla possess the entire synthesis pathway [89, 90]. Most bacteria and phytoplankton are auxotrophic for  $B_{12}$  and are thus dependent on the few taxa capable of  $B_{12}$  production. Class II RTPR viroplankton populations must infect a host that can either synthesize  $B_{12}$  or assimilate and subsequently modify a  $B_{12}$  precursor containing the corrin ring. Moreover,  $B_{12}$  has several chemical analogs depending on the combination of ligands at the  $\alpha$  and  $\beta$  sites of the corrin ring [91, 92].

There is a growing recognition that selectivity in  $B_{12}$  analogs may drive interaction networks within microbial communities as cobamide specificity is commonly observed in microorganisms [93]. For example, pseudocobalamin, produced by cyanobacteria such as *Prochlorococcus* (the dominant bacterioplankton taxa observed in this study), is poorly utilized by  $B_{12}$  auxotrophic microalgae [92, 94]. Specificity in  $B_{12}$  requirements among bacterioplankton taxa may explain the separation observed

between bacterioplankton populations associating with either SAR11\_clade\_II populations or *Prochlorococcus*\_MIT9313 (Groups 1 and 2, respectively, in Supplementary Fig. S11). Positive association patterns between viroplankton Class II RTPR phyto-type populations and bacterioplankton 16S ASV populations reflected the split seen between SAR11 and *Prochlorococcus* associations. Class II RTPR viroplankton clades 5 and 6 associated with SAR11\_clade\_II and 14 other bacterioplankton populations (Fig. 6), three of which (Magnetospiraceae, Rhodobacteraceae, Planctomarina) may be capable of  $B_{12}$  production [92]. Eleven of these bacterioplankton populations are associated with SAR11\_clade\_II (Supplementary Fig. S11, Group 1). Similarly, Class II RTPR viroplankton clades 1 and 3 associated with *Prochlorococcus*\_MIT9313, and six other bacterioplankton populations that also demonstrated associations with *Prochlorococcus* (Supplementary Fig. S10, Group 2). Three of these seven (AEGEAN\_169\_marine\_group\_1, AEGEAN\_169\_marine\_group\_2, and *Prochlorococcus*\_MIT9313) occur in an order, family, or genera containing species capable of  $B_{12}$  production (Fig. 6).

Time- and depth-resolved observations of Class II RTPR viroplankton clades 1, 3, and 5 (Supplementary Figs. S12 and S13) indicate that changes in the abundance of these viroplankton populations followed those of their associated bacterioplankton populations throughout the euphotic zone. It seems unlikely that the parallels in association patterns observed between bacterioplankton populations (Supplementary Fig. S11) and those observed between Class II RTPR viroplankton and bacterioplankton populations (Fig. 6) are happenstance. It is likely that these positive virio- to bacterioplankton association patterns reflect the  $B_{12}$ -driven association patterns among their bacterioplankton hosts [93]. A few heterotrophic bacterial taxa and *Prochlorococcus* were likely the principal  $B_{12}$ -producing populations in the euphotic surface waters of the Azores. Although Thaumarchaeota

are also important oceanic  $B_{12}$  producers, metagenomic biogeographic surveys indicate that these archaeal populations predominate at high latitudes or in mesopelagic waters [90].

It has been hypothesized that viroplankton influence oceanic primary and secondary productivity by controlling the flux of limiting trace nutrients, such as iron [95], through cell lysis. The common presence and dynamic nature of Class II RTPR viroplankton populations now places viral–host interactions as another important component within the vitamin traffic pathways of this trace nutrient in the oligotrophic ocean [96]. While technically  $B_{12}$  is not a nutrient as it is not incorporated into cellular biomass, it displays oceanic distribution patterns reminiscent of nutrients such as nitrogen. Within surface waters of the oligotrophic oceanic gyres,  $B_{12}$  occurs in vanishingly low concentrations of less than 1.75 picomolar [92, 97]. However, deeper waters typically demonstrate higher  $B_{12}$  concentrations [98]. Depth-resolved dissolved  $B_{12}$  concentrations occurring at our study site were likely highest below 25 m, similar to those recorded for in April and August at the western-most station of a recent study of eastern north Atlantic central waters off the Iberian peninsula [97]. While Class II RTPR viroplankton populations did show differences in their depth and time distribution (Fig. 8, Supplementary Figs. S12 and S13), the abundance of these populations as a fraction of overall viroplankton abundance could not be discerned using an amplicon approach. Nevertheless, as Class II RTPR viral populations depend on  $B_{12}$  for a steady supply of deoxyribonucleotides, it is possible that these viroplankton regulate or enhance  $B_{12}$  uptake or production during infection. Subsequently, lysis by Class II RTPR-utilizing viroplankton could increase the flux of  $B_{12}$  pools disproportionately over viroplankton populations that do not utilize  $B_{12}$ -dependent proteins. New quantitative approaches for observations of viral–host interactions based on detecting single viral genes using polonies [99–101] could provide data on the in situ frequency and production rate of Class II RTPR viroplankton populations, as has been recently demonstrated for pelagic cyanophage populations [102].

Unlike Class I and dimeric Class II RNRs that use ribonucleotide diphosphate (NDP) substrates, Class II RTPR uses ribonucleotidetriphosphate (NTP) substrates. This difference in substrates may correlate with significant differences in viral population ecology. Viroplankton populations utilizing a Class I or dimeric Class II RNR would be dependent on NDP production from RNA digestion [103] as well as an NDP kinase for producing dNTPs [104]. In contrast, Class II RTPR viroplankton populations have direct and immediate access to the cellular pool of NTPs (which is substantially larger than the NDP pool [105]) for DNA synthesis. These biochemical features of Class II RTPR would favor a rapid lytic cycle. It is possible that Class II RTPR-carrying viroplankton are highly lytic viruses specifically infecting actively growing bacterioplankton populations containing large intercellular pools of rNTP substrates.

However, there are possible costs to a Class II RTPR strategy in terms of DNA synthesis regulation. Improved control of dNTP substrate levels for DNA synthesis and allosteric regulation of RNR enzyme activity have been proposed as possible advantages of NDP versus NTP reduction [85]. Thus, Class II RTPR-carrying viruses may pay a cost for faster DNA synthesis with lower fidelity as imbalances or excesses in dNTP levels from runaway Class II RTPR activity can increase mutagenesis rates [104, 106]. The fact that Class II RTPRs are common within the viroplankton [9, 34, 81] and are proportionally more represented within viruses than within cellular genomes (Supplementary Fig. S1) indicates that the possible fitness costs of Class II RTPR do not outweigh the benefits of utilizing this gene for dNTP synthesis. Nevertheless, it is possible that the frequency of Class II RTPR-carrying viruses may depend on environmental context as the majority of known viruses carry a Class I or Class II dimeric RNR (Supplementary Fig. S1).

### Class II RTPR viroplankton phylotypes demonstrated dynamic responses to ecosystem change

At the community and population scale, Class II RTPR-carrying viroplankton demonstrated dynamic behavior throughout the four month study, responding to seasonal and oceanographic changes influencing the activity of co-occurring phyto- and bacterioplankton communities. In this sense, Class II RTPR viroplankton populations demonstrated community dynamic behavior reminiscent of that seen from studies utilizing amplicons of structural genes such as T4 major capsid protein [28, 30], portal protein [24, 107], and terminase [25–27]. A unique aspect of this study was that primer sets targeting viral Class II RTPR genes were used for amplifying Class II RTPR from both the viroplankton (0.02–0.22  $\mu\text{m}$ ) and bacterioplankton (>0.22  $\mu\text{m}$ ) fractions. Prior investigations identifying and tracking T4-like cyanophage populations within virome and metagenome assemblies using the gp43 gene for PolB (DNA polymerase) found that viromes contained seasonal and persistent populations whereas metagenomes contained mostly sporadically observed populations [108]. These observed differences were attributed to the fact that free viruses (represented within virome data) represent prior active infections; whereas viruses within cells (represented within metagenome data) are undergoing active infection. The majority of Class II RTPR OTUs observed in both fractions showed a greater abundance in the viroplankton fraction (Fig. 7) indicating host populations undergoing active lysis. These OTUs had the highest cl abundance of all Class II RTPR viroplankton populations observed in the study. This observation makes sense as each viral particle released contains a Class II RTPR gene copy and dozens of viral particles are released for each lysed cell. Class II RTPR OTUs observed solely within either the viro- or bacterioplankton fraction were less frequent (8% viroplankton only; 16% bacterioplankton only) and typically showed low cl abundance. Viroplankton-only Class II RTPR OTUs likely represented viral populations either no longer actively infecting hosts or somehow displaced from their host populations. Bacterioplankton-only Class II RTPR OTUs may have represented lysogenic viruses existing as prophages, free viral genomic DNA within the >0.22  $\mu\text{m}$  size fraction, some lytic viral populations yet to lyse infected cells, or inadvertent amplification of a bacterial Class II RTPR gene. Distinguishing between these four possibilities would be difficult; nevertheless, it is clear that the majority of observed Class II RTPR OTUs (84%) occurred within the viroplankton.

Individual Class II RTPR viroplankton populations were actively infecting, replicating, and subsequently disappearing from the pelagic euphotic zone. Using an amplicon approach provided the advantage of deeply sampling Class II RTPR viroplankton; it was less clear how these data should be partitioned into ecologically meaningful populations. Community and population ecology analyses conducted using either a heuristic approach (binning amplicons into 98% amino acid identity OTUs) or an evolutionary approach (defining major phylogenetic clades of Class II RTPR amplicons) demonstrated the superiority of the evolutionary approach. Community beta diversity analyses based on cl-transformed OTU or clade abundance showed that the evolutionary approach (clades, Fig. 5B and Supplementary Fig. S6B) captured a greater proportion of variability in viroplankton communities than the heuristic approach (98% OTUs Supplementary Fig. S7C, D). The evolutionary approach lumped Class II RTPR viroplankton populations into 17 clades as opposed to thousands of 98% OTUs, likely providing a greater signal to noise ratio for observing viroplankton ecological behavior. Indeed, viroplankton evolutionary relationships gleaned from Class II RTPR phylogeny demonstrated a significant connection with the ecological behavior of viroplankton populations on the August 5th sampling date (Fig. 8). The Class II RTPR genes carried within each clade population likely occurred within a diversity of genomic backgrounds. Thus, it was surprising that this single gene showed a

statistically significant positive relationship between evolutionary distance (Class II RTPR phylogenetic tree) and patterns of cl abundance within the bacterioplankton fraction (Aitchison distance) (Supplementary Fig. S15). The significant correlation of cophenetic distance (Class II RTPR phylogeny) and ecological behavior (Aitchison distance) of Class II RTPR OTUs identified within the bacterioplankton and not within free viruses is interesting as these populations represented those Class II RTPR viral populations undergoing active replication.

Examinations of connections between gene-based viral phylogeny and viral ecology have been limited, in large part due to the widely acknowledged mosaicism within viral genomes [109]. Nevertheless, when considering the influence of enzyme biochemistry or the presence/absence of particular auxiliary metabolic genes [110], it is clear that single genes have a dramatic influence on a virus' ecology. Single-gene amplicon approaches provide not only an evolutionary framework for understanding viral ecology, but also have practical advantages over metagenomic approaches in terms of cost, depth and breadth of sampling, and computational complexity [111]. Prior work examining genetic connections to phage host range phenotypes found that while whole genome networks provided better predictive capability, single-gene groups could also show strong connections [112]. Interestingly, DNA replication and phage replication were among those Rapid Annotation Subsystem gene groups [113] showing a stronger predictive ability for host range. This might not be surprising when considering that phage genome and replication processes may require intimate interaction with host anabolic systems. Examining possible connections between single genes and whole genomes and other important phage phenotypes, especially those describing infection dynamics, should ultimately improve the predictive ability of metagenome data for ecosystem modeling of phage impacts on ecosystem processes.

## DATA AVAILABILITY

The Class II RTPR and 16S rRNA gene sequence datasets generated and analyzed during the current study have been deposited with links to BioProject accession number PRJNA842570 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>). Sequences and Additional Files generated and analyzed during the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.8183150>.

## REFERENCES

- McClain CR, Signorini SR, Christian JR. Subtropical gyre variability observed by ocean-color satellites. *Deep Sea Res. Part 2 Top Stud Oceanogr.* 2004;51:281–301.
- Huang RX, Russell S. Ventilation of the subtropical North Pacific. *J Phys Oceanogr.* 1994;24:2589–605.
- Sherr E, Sherr B. Understanding roles of microbes in marine pelagic food webs: a brief history. in *Microbial Ecology of the Oceans*. Hoboken, New Jersey. 2008. pp. 27–44.
- Goldman JC. Conceptual role for microaggregates in pelagic waters. *Bull Mar Sci.* 1984;35:462–76.
- Díez B, Pedrós-Alió C, Massana R. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol.* 2001;67:2932.
- Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience.* 1999;49:781–8.
- Ban N, Alder J. How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada. *Aquat Conserv.* 2008;18:55–85.
- Li D, Daler D. Ocean pollution from land-based sources: East China Sea, China. *Ambio.* 2004;33:107–13.
- Sakowski EG, Munsell EV, Hyatt M, Kress W, Williamson SJ, Nasko DJ, et al. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. *Proc Natl Acad Sci USA.* 2014;111:15786–91.
- Harrison AO, Moore RM, Polson SW, Wommack KE. Reannotation of the ribonucleotide reductase in a cyanophage reveals life history strategies within the viroplankton. *Front Microbiol.* 2019;10:134.
- Lara E, Vaqué D, Sà EL, Boras JA, Gomes A, Borrull E, et al. Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci Adv.* 2017;3:e1602565.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature.* 1999;399:541–8.
- Suttle CA. Viruses in the sea. *Nature.* 2005;437:356–61.
- Wommack KE, Colwell RR. Viroplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev.* 2000;64:69–114.
- Jiang SC, Paul JH. Gene transfer by transduction in the marine environment. *Appl Environ Microbiol.* 1998;64:2780–7.
- Schwartz DA, Lindell D. Genetic hurdles limit the arms race between *Prochlorococcus* and the T7-like podoviruses infecting them. *ISME J.* 2017;11:1836–51.
- Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000;405:299–304.
- Rosenwasser S, Ziv C, van Creveld SG, Vardi A. Virocell metabolism: metabolic innovations during host–virus interactions in the *Ocean*. *Trends Microbiol.* 2016;24:821–32.
- Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology.* 2016;499:219–29.
- Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA.* 2011;108:E757–64.
- Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol.* 1991;173:697–703.
- Schmidt HF, Sakowski EG, Williamson SJ, Polson SW, Wommack KE. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine viroplankton. *ISME J.* 2014;8:103–14.
- Brussaard CPD, Short SM, Frederickson CM, Suttle CA. Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). *Appl Environ Microbiol.* 2004;70:3700–5.
- Zhong Y, Chen F, Wilhelm SW, Poorvin L, Hodson RE. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. *Appl Environ Microbiol.* 2002;68:1576–84.
- Short CM, Suttle CA. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl Environ Microbiol.* 2005;71:480–6.
- Marston MF, Sallee JL. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol.* 2003;69:4639–47.
- Goldin S, Hulata Y, Baran N, Lindell D. Quantification of T4-Like and T7-like cyanophages using the polony method show they are significant members of the viroplankton in the North Pacific subtropical gyre. *Front Microbiol.* 2020;11:1210.
- Comeau AM, Krisch HM. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol Biol Evol.* 2008;25:1321–32.
- Needham DM, Sachdeva R, Fuhrman JA. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* 2017;11:1614–29.
- Liu L, Cai L, Zhang, R. Co-existence of freshwater and marine T4-like myoviruses in a typical subtropical estuary. *FEMS Microbiol Ecol.* 2017;93:fix119.
- Lundin D, Gribaldo S, Torrents E, Sjöberg B-M, Poole AM. Ribonucleotide reduction - horizontal transfer of a required function spans all three domains. *BMC Evol Biol.* 2010;10:383.
- Nordlund P, Reichard P. Ribonucleotide reductases. *Annu Rev Biochem.* 2006;75:681–706.
- Herrick J, Sclavi B. Ribonucleotide reductase and the regulation of DNA replication: an old story and an ancient heritage. *Mol Microbiol.* 2007;63:22–34.
- Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol Biol.* 2013;13:33.
- Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* 2005;3:e144.
- Madigan MT, Clark DP, Stahl D, Martinko JM. *Brock biology of microorganisms*, 13th ed. Upper Saddle River: Prentice Hall; 2003. p. 1019.
- Chen F, Lu J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol.* 2002;68:2589–94.
- Rohwer F, Segall A, Steward G. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol Oceanogr.* 2000;45:408–18.

39. Lundin D, Torrents E, Poole AM & Sjöberg B-M. RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of mis-annotation in sequences deposited to Genbank. *BMC Genom.* 2009;10. <https://doi.org/10.1186/1471-2164-10-589>.
40. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49:D480–9.
41. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol.* 2014;10:1–12.
42. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017;5:27.
43. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 2017;8:2224.
44. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. *Elife.* 2019;8:e46923.
45. Bian G, Gloor GB, Gong A, Jia C, Zhang W, Hu J, et al. The gut microbiota of healthy aged Chinese is similar to that of the healthy young. *mSphere.* 2017;2:e00327-17.
46. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* 2014;2:15.
47. Publication Reports - EcosystemOverview\_Azores\_2019.pdf. <https://www.ices.dk/sites/pub/Publication%20Reports/Forms/DispForm.aspx?ID=36442>, <https://doi.org/10.17895/ices.advice.5753>.
48. Bashmachnikov I, Belonenko TV, Koldunov AV. Intra-annual and interannual non-stationary cycles of chlorophyll concentration in the Northeast Atlantic. *Remote Sens. Environ.* 2013;137:55–68.
49. Aristegui J, Mendonça A, Vilas JC, Espino M, Polo I, Montero MF, et al. Plankton metabolic balance at two North Atlantic seamounts. *Deep Sea Res Part 2 Top Stud Oceanogr.* 2009;56:2646–55.
50. Mendonça A, Aristegui J, Vilas JC, Montero MF, Ojeda A, Espino M, et al. Is there a seamount effect on microbial community structure and biomass? The case study of Seine and Sedlo seamounts (northeast Atlantic). *PLoS ONE.* 2012;7:e29526.
51. True colors of oceanography: guidelines for effective and accurate colormap selection | Oceanography. <https://doi.org/10.5670/oceanog.2016.66>.
52. Schlitzer R. Interactive analysis and visualization of geoscience data with Ocean Data View. *Comput Geosci.* 2002;28:1211–8.
53. Sotillo MG, Cailleau S, Lorente P, Levier B, Aznar R, Refray G, et al. The MyOcean IBI Ocean Forecast and Reanalysis Systems: operational products and roadmap to the future Copernicus Service. *J Oper Oceanogr.* 2015;8:63–79.
54. Baith K, Lindsay R, Fu G, McClain CR. Data analysis system developed for ocean color satellite sensors. *Eos Trans AGU.* 2001;82:202.
55. Marie D, Partensky F, Vaulot D, Brussaard C. Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr Protoc Cytom.* 2001;Chapter 11:Unit 11.11.
56. Iker BC, Kitajima M, Gerba CP. Extraction and purification of viral nucleic acids from environmental samples. In: Micic M, editor. *Sample preparation techniques for soil, plant, and animal samples.* New York: Springer; 2016. p. 315–24.
57. Mueller JA, Cullley AI, Steward GF. Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. *Appl Environ Microbiol.* 2014;80:3930–42.
58. Fuhrman JA, Comeau DE, Hagström A, Chan AM. Extraction from natural planktonic microorganisms of DNA suitable for molecular biological studies. *Appl Environ Microbiol.* 1988;54:1426–9.
59. Fadrosch DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome.* 2014;2:6.
60. Moore RM, Harrison AO, Nasko DJ, Chopyk J, Cebeci M, Ferrell BD, et al. PASV: Automatic protein partitioning and validation using conserved residues. *bioRxiv.* 2021 <https://doi.org/10.1101/2021.01.20.427478>.
61. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26:2460–1.
62. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5:1–10.
63. Moore RM, Harrison AO, McAllister SM, Polson SW, Wommack KE. Iroki: automatic customization and visualization of phylogenetic trees. *PeerJ.* 2020;8:e8584.
64. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7.
65. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581.
66. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2012;41:D590–6.
67. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967;27:209–20.
68. Willis AD & Martin, BD DivNet: Estimating diversity in networked communities. *bioRxiv.* 2018. <https://doi.org/10.1101/305045>.
69. Wickham, H *ggplot2: Elegant graphics for data analysis.* Springer; 2016.
70. Willis A, Bunge J. Estimating diversity via frequency ratios. *Biometrics.* 2015;71:1042–9.
71. Kang S, Rodrigues JLM, Ng JP, Gentry TJ. Hill number as a bacterial diversity measure framework with high-throughput sequence data. *Sci Rep.* 2016;6:38263.
72. van den Boogaart KG, Tolosana-Delgado R, Bren M. compositions: compositional data analysis. *Austral Ecol.* 2018;26:32–46.
73. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001;26:32–46.
74. Quinn TP, Richardson MF, Lovell D, Crowley TM. propr: An R-package for identifying proportionally abundant features using compositional data analysis. *Sci Rep.* 2017;7:16252.
75. Erb I, Notredame C. How should we measure proportionality on relative gene expression data? *Theory Biosci.* 2016;135:21–36.
76. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9.
77. Palarea-Albaladejo J, Martín-Fernández JA. zCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics Intellig Lab Syst.* 2015;143:85–96.
78. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021.
79. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol.* 2016;62:692–703.
80. Alberti A, Poulain J, Engelen S, Labadie K, Romac S, Ferrera I, et al. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci Data.* 2017;4:170093.
81. Wommack KE, Nasko DJ, Chopyk J, Sakowski EG. Counts and sequences, observations that continue to change our understanding of viruses in nature. *J Microbiol.* 2015;53:181–92. <https://doi.org/10.1007/s12275-015-0068-6>.
82. Bidle KD. Elucidating marine virus ecology through a unified heartbeat. *Proc Natl Acad Sci USA.* 2014;111:15606–7.
83. Stubbe J. Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr Opin Struct Biol.* 2000;10:731–6.
84. Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie.* 2005;87:793–803.
85. Lundin D, Berggren G, Logan DT, Sjöberg B-M. The origin and evolution of ribonucleotide reduction. *Life.* 2015;5:604–36.
86. Martínez-Hernández F, Fornas O, Llesmas Gomez M, Bolduc B, De la Cruz Peña MJ, Martínez JM, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun.* 2017;8:15892.
87. Sakowski EG, Arora-Williams K, Tian F, Zayed AA, Zablocki O, Sullivan MB, et al. Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR. *Nat Microbiol.* 2021;6:630–42.
88. Torrents E. Ribonucleotide reductases: essential enzymes for bacterial life. *Front Cell Infect Microbiol.* 2014;4. <https://doi.org/10.3389/fcimb.2014.00052>.
89. Sañudo-Wilhelmy SA, Gómez-Consarnau L, Suffridge C & Webb EA. The role of B vitamins in marine biogeochemistry. *Ann Rev Mar Sci.* 2014;6:339–67. <https://doi.org/10.1146/annurev-marine-120710-100912>.
90. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J.* 2015;9:461–71.
91. Grossman A. Nutrient acquisition: the generation of bioactive vitamin B12 by microalgae. *Curr Biol.* 2016;26:R319–21.
92. Heal KR, Qin W, Ribalet F, Bertagnolli AD, Coyote-Maestas W, Hmelo LR, et al. Two distinct pools of B 12 analogs reveal community interdependencies in the ocean. *Proc Natl Acad Sci USA.* 2017;114:364–9. <https://doi.org/10.1073/pnas.1608462114>.
93. Sokolovskaya OM, Shelton AN, Taga ME. Sharing vitamins: cobamides unveil microbial interactions. *Science.* 2020;369:eaba0165.
94. Helliwell KE, Lawrence AD, Holzer A, Kudahl UJ, Sasso S, Kräutler B, et al. Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. *Curr Biol.* 2016;26:999–1008. <https://doi.org/10.1016/j.cub.2016.02.041>.
95. Poorvin L, Rinta-Kanto JM, Hutchins DA, Wilhelm SW. Viral release of iron and its bioavailability to marine plankton. *Limnol Oceanogr.* 2004;49:1734–41.
96. Giovannoni SJ. Vitamins in the sea. *Proc Natl Acad Sci USA.* 2012;109:13888–9.
97. Joglar V, Álvarez-Salgado XA, Gago-Martínez A, Leao JM, Pérez-Martínez C, Pontiller B, et al. Cobalamin and microbial plankton dynamics along a coastal to



- offshore transect in the Eastern North Atlantic Ocean. *Environ Microbiol.* 2021;23:1559–83. <https://doi.org/10.1111/1462-2920.15367>.
98. Sañudo-Wilhelmy SA, Cutter LS, Durazo R, Smail EA, Gómez-Consarnau L, Webb EA, et al. Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc Natl Acad Sci USA.* 2012;109:14041–5.
  99. Mitra RD, Shendure J, Olejnik J, Edyta-Krzyszowska-Olejnik, Church GM. Fluorescent in situ sequencing on polymerase colonies. *Anal Biochem.* 2003;320:55–65.
  100. Mruwat N, Carlson MC, Goldin S, Ribalet F, Kirzner S, Hulata Y, et al. A single-cell polony method reveals low levels of infected *Prochlorococcus* in oligotrophic waters despite high cyanophage abundances. *ISME J.* 2021;15:41–54.
  101. Baran N, Goldin S, Maidanik I, Lindell D. Quantification of diverse virus populations in the environment using the polony method. *Nat Microbiol.* 2018;3:62–72.
  102. Carlson MCG, Ribalet F, Maidanik I, Durham BP, Hulata Y, Ferrón S, et al. Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat Microbiol.* 2022;7:570–80.
  103. Zuo Y, Deutscher MP. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 2001;29:1017–26.
  104. Mathews CK. Deoxyribonucleotides as genetic and metabolic regulators. *FASEB J.* 2014;28:3832–40.
  105. Varik V, Oliveira SRA, Haurlyuk V, Tenson T. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Sci Rep.* 2017;7:11022.
  106. Wheeler LJ, Rajagopal I, Mathews CK. Stimulation of mutagenesis by proportional deoxyribonucleoside triphosphate accumulation in *Escherichia coli*. *DNA Repair.* 2005;4:1450–6.
  107. Dedeo CL, Cingolani G, Teschke CM. Portal protein: the orchestrator of capsid assembly for the dsDNA tailed bacteriophages and herpesviruses. *Annu Rev Virol.* 2019;6:141–60.
  108. Dart E, Fuhrman JA, Ahlgren NA. Diverse marine T4-like cyanophage communities are primarily comprised of low-abundance species including species with distinct seasonal, persistent, occasional, or sporadic dynamics. *Viruses.* 2023;15:581.
  109. Hatfull GF. Bacteriophage genomics. *Curr Opin Microbiol.* 2008;11:447–53.
  110. Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol.* 2016;31:161–8.
  111. Adriaenssens EM, Cowan DA. Using signature genes as tools to assess environmental viral ecology and diversity. *Appl Environ Microbiol.* 2014;80:4470–80.
  112. Shapiro JW, Putonti C. Gene co-occurrence networks reflect bacteriophage ecology and evolution. *MBio.* 2018;9:e01870–17.
  113. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014;42:D206–14.

## ACKNOWLEDGEMENTS

We would like to thank, by order of involvement, Catharina Pieper, Ricardo Fernandes, Christien Laber, Prasanna Joglekar, Joana Botelho, Ana Pavon, Clara M. Loureiro, and Marília Olio for the help collecting and processing the samples. We would like to thank Dr. Daniel Nasko for input on library preparation and statistical analysis. We would like to acknowledge the skipper Renato Bettencourt, and all laboratory and facility management personnel from Departamento de Oceanografia e P, OKEANOS Institute of Marine Research, and IMAR-Instituto do Mar from the University of the Azores, DMCS-RU and DBI-UD.

## AUTHOR CONTRIBUTIONS

GP designed the study, performed processing of water samples and DNA extraction following water samples collection and part of the Class II RTPR library preparation. LW conducted the majority of the Class II RTPR library preparation and all of the 16S rDNA library preparation, performed all of the bioinformatics and biostatistical analyses following sequence processing, and wrote and edited the manuscript. RM, CL, AH, JD, BF, AM, KB, SP, JN and EW contributed to the study design, protocol development, data analysis and interpretation, and manuscript preparation. All authors read, edited, and approved the final manuscript.

## FUNDING

This research was supported by the Estagiar-L fellowship; EEA grant funded BIOMETOREzores (PT02\_Aviso2\_0001) project; the mobility program FLAD-UA Crossing the Atlantic; Chinese Scholarship Council (CSC), and National Science Foundation grant number 1736030. Support from the University of Delaware Center for Bioinformatics and Computational Biology (CBCB) Core Facility, the University of Delaware Sequencing and Genotyping Center, and use of the BIOMIX compute cluster was made possible through funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43705-023-00306-9>.

**Correspondence** and requests for materials should be addressed to K. Eric Wommack.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023